

ŽILINSKÁ UNIVERZITA V ŽILINE
Fakulta riadenia
a informatiky

Fakulta riadenia a informatiky

Dátová analytika v športovom manažmente

Diplomová práca

BC. ALEXANDER SÝKORA

Študijný program: Informačný manažment

Študijný odbor: Ekónómia a manažment

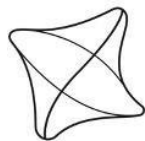
Školiace pracovisko: Žilinská univerzita v Žiline

Vedúci diplomovej práce: Ing. Lucia Pančíková, PhD.

Ministerské číslo práce: 28360020232043

Registračné číslo práce: 2192/2022

Žilina 2023



ŽILINSKÁ UNIVERZITA V ŽILINE
Fakulta riadenia
a informatiky

DIPLOMOVÁ PRÁCA

Bc. Alexander Sýkora

Dátová analytika v športovom manažmente
Data analytics in sports management

Študijný odbor: ekonómia a manažment

Študijný program: informačný manažment

Stupeň kvalifikácie: inžinier (Ing.)

Dátum zadania práce: 31.10.2022

Dátum odovzdania práce: 19.4.2023

Žilina, 2023

ZADANIE TÉMY DIPLOMOVEJ PRÁCE.

Študijný program: Informačný manažment

Meno a priezvisko

Alexander Sýkora

Osobné číslo

560781

Názov práce v slovenskom aj anglickom jazyku

Dátová analytika v športovom manažmente

Data analytics in sports management

Zadanie úlohy, ciele, pokyny pre vypracovanie

(Ak je málo miesta, použite opačnú stranu)

Cieľ diplomovej práce:

Na základe teoretických východísk v problematike športového manažmentu a dátovej analytiky vypracovať návrh na zlepšenie súčasného stavu v špecifikovanej oblasti, s implementáciou modelového riešenia.

Obsah:

1. Teoretické východiská a súčasný stav riešenej problematiky so zohľadnením dátovej analytiky.
2. Metodologická časť práce.
3. Súčasný stav riešenia s implementáciou dátovej analytiky v športovom odvetví.
4. Analýza dátových zdrojov a ich väzieb, výber modelovacích prostriedkov.
5. Návrh pre zlepšenie súčasného stavu v špecifikovanej oblasti športového manažmentu s využitím modelového prístupu. Modelové riešenie predpokladá využitie Open Sourceového programovacieho prostredia.
6. Ekonomické a mimoeconomické vyhodnotenie, prínosy návrhu z hľadiska vybranej oblasti športového manažmentu.

Pokyny pre vypracovanie: Diplomovú prácu treba vypracovať v súlade s riadiacimi aktivitami vybraného subjektu, je potrebné pracovať s reálnymi údajmi. Pri analýzach a vypracovaní práce treba postupovať podľa interných smerníc fakulty.

Meno a pracovisko vedúceho DP: Ing. Lucia Pančíková, PhD., KMME, ŽU

Meno a pracovisko tútora DP:

31 OKT. 2022

vedúci katedry
(dátum a podpis)

Zadanie zaregistrované dňa 31.10.2022 pod číslom 2192/2022 podpis

Čestné vyhlásenie

Vyhlasujem, že som zadanú diplomovú prácu vypracoval samostatne, pod odborným vedením vedúceho práce/školiťa a používal som len literatúru uvedenú v práci.

Žilina 19. apríl 2023

podpis

Pod'akovanie

Ďakujem Ing. Lucíí Pančfkovej, PhD. za odborné vedie, ústretovú komunikáciu, hodnotné rady a pripomienky počas vypracovania diplomovej práce. Vďaka patrí aj doc. Ing. Michalovi Varmusovi, PhD. za sprostredkovanie kontaktov a Slovenskému futbalovému zväzu za sprístupnenie informácií a aktívne komunikovanie vedomostí a znalostí.

Abstrakt

SÝKORA, Alexander: *Dátová analytika v športovom manažmente*. [Diplomová práca]. – Žilinská Univerzita. Fakulta riadenia a informatiky; Katedra makro a mikroekonomiky. – Vedúca práce: Ing. Lucia Pančíková, PhD. Stupeň odbornej kvalifikácie: inžinier. - Žilina: FRI UNIZA, 2023. 80 s.

Diplomová práca sa zaoberá využitím dátovej analýzy v športovom manažmente. Cieľom práce je na základe teoretických východísk v oblasti športového manažmentu a dátovej analytiky vypracovať návrh na zlepšenie súčasného stavu v špecifickej problematike, s implementáciou modelového riešenia. Diplomová práca je rozdelená na tri hlavné časti. V prvej časti autor poskytuje teoretické východiská športového manažmentu a dátovej analytiky. Prvá časť tiež rozoberá techniky a metódy klasifikácie výstupu a laženia dát. V druhej časti autor definuje konkrétny problém z problematiky, ktorý sa v práci analyzuje viacerými metódami, technikami a postupmi. Primárnym zdrojom analýzy sú dáta poskytnuté SFZ. V závere druhej časti autor vytvorí modely a zhodnotí ich použiteľnosť. Tretia časť diplomovej práce prezentuje návrhy na zlepšenie súčasného stavu vo vybranom probléme a špecifikuje limitácie návrhu. V rámci tejto časti práce sa prezentujú ekonomické a mimoekonomické vyhodnotenia návrhov.

Kľúčové slová: športový manažment, modelovanie v športe, dolovanie dát z databázy, dátová analytika

Abstract

SÝKORA, Alexander: *Data analytics in sports management*. [Master thesis]. – University of Zilina. Faculty of Management Science and Informatics; Department of Macro and Microeconomics. – Tutor: Ing. Lucia Pančíková, PhD. Qualification level: Master of Science. – Žilina: FRI UNIZA, 2023. 80 p.

The diploma thesis deals with data analysis in sports management. The thesis aims to develop a proposal for improving the current situation based on theoretical principles in the field of sports management and data analysis in a specific area, with the implementation of a model solution. The diploma thesis is divided into three main parts. In the first part, the author provides the theoretical basis of sports management and data analysis. Additionally, this part of the thesis discusses the techniques and methods of output classification and data mining. In the second part of the thesis, the author defines a specific problem within the area studied, which is then analysed using several methods, techniques, and procedures. The primary source of analysis is the data provided by Slovak Football Association. The end of second part of the thesis contains classification models created by the author and their usability evaluation. The third part of the thesis presents proposals to improve the current state of selected problems and specifies the limitations of the proposals. This part of the diploma thesis also includes the evaluation of economic and non-economic aspects of proposals.

Keywords: sports management, modelling in sports, data mining in databases, data analytics

Obsah	13
1 Teoretické východiská	14
1.1 Športový manažment	14
1.1.1 Dátová analýza v športovom manažmente	14
1.2 Ťaženie dát	15
1.2.1 Dolovanie vedomostí z databázy	17
1.2.2 Využitie ťaženia dát pre podporu rozhodovania	18
1.3 Dátové modelovanie	19
1.3.1 Učenie s dozorom	19
1.3.2 Klasifikačné modely	20
1.4 Využitie modelovania na odchod z organizovaného futbalu	24
2 Metodológia práce	26
2.1 Cieľ práce	26
2.2 Metódy, techniky a postupy	26
3 Analytická časť	28
3.1 Definovanie problému	28
3.1.1 Odchod z organizovaného športu po dovišení 15. roku	28
3.2 Faktory vplývajúce na odchod z organizovaného športu	29
3.2.1 Záujem o šport	29
3.2.2 Náročnosť súťaže	31
3.2.3 Trénovanie	32
3.2.4 Vyhorenie	33
3.2.5 Čas	34
3.2.6 Ostatné faktory	34
3.2.7 Zhrnutie faktorov vplývajúcich na odchod	35
3.3 Spracovanie dát	36
3.3.1 Príprava dát na analýzu	36
3.3.2 Vytvorenie dátového skladu	38
3.4 Dolovania dát poskytnutých SFZ	41
3.4.1 Zhrnutie dolovania dát SFZ	45
3.5 Analýza dát z reportov „Šport v číslach“	46
3.5.1 Zhrnutie analýza reportov Šport v číslach	49
3.6 Modelovanie odchodu mládeže z organizovaného futbalu	49
3.6.1 Logistická regresia	52
3.6.2 Rozhodovací strom	53

Obsah	
3.6.3	Náhodné lesy 54
3.6.4	Metóda podporných vektorov 54
3.6.5	Zhrnutie modelovania odchodu mládeže z organizovaného futbalu 56
4	Návrhová časť 57
4.1	Potrebné dáta na vysvetlenie problému 57
4.1.1	Spôsob získavania dát 58
4.1.2	Scenáre návrhu získania dát 59
4.2	Zvýšenie dimenzií odchodu zo SFZ 60
4.2.1	Vyššia veková kategória 61
4.2.2	Prestup do iného zväzu 62
4.2.3	Zmena športu 62
4.2.4	Skončenie so športom 63
4.2.5	Scenáre návrhu zvýšenia dimenzií výstupu 64
5	Prínosy a Vyhodnotenie návrhov 65
5.1	Mimoekonomické vyhodnotenie 65
5.2	Ekonomické vyhodnotenie 67
	Záver 69

Zoznam obrázkov

Obrázok 1.2.1 Cyklus ťaženia dát CRISP-DM	16
Obrázok 1.3.1 Logistická regresia	21
Obrázok 1.3.2 Naivný Bayesov klasifikátor	22
Obrázok 1.3.3 Rozhodovací strom	22
Obrázok 1.3.4 Náhodný les	23
Obrázok 1.3.5 Podpora vektorových strojov: Lineárna a Nelineárna	24
Obrázok 1.4.1 Výsledok štúdie Van Yperen a kolektívu 2022.....	25
Obrázok 3.1.1 Vývoj vekových kategórií vo výskume ICOACHKIDS – futbal	28
Obrázok 3.2.1 Myšlienková mapa faktorov na základe analýzy literatúry	29
Obrázok 3.3.1 Diagram zdrojov, nástrojov a fáz spracovania dát.....	36
Obrázok 3.3.2 Ukážka skriptu dátového kanálu v MongoDB Compas	38
Obrázok 3.3.3 Ukážka výstupu dátového kanálu v MongoDB Compas	38
Obrázok 3.3.4 Vzor ERD diagramu dátového skladu	39
Obrázok 3.3.5 Vzor výstupu agregácie SFZ zálohy v MySQL.....	41
Obrázok 3.5.1 Počet futbalistov na ZŠ podľa krajov porovnanie roku 2015 a 2021	48
Obrázok 3.5.2 Počet futbalistov na SŠŠ podľa krajov porovnanie roku 2015 a 2021.....	48
Obrázok 3.6.1 Korelačná matica tréningovej množiny	51
Obrázok 3.6.2 Výstup logistická regresia – koeficienty	52
Obrázok 3.6.3 Výsledný rozhodovací strom.....	53
Obrázok 3.6.4 Vývoj chýb počas vytvárania jednotlivých stromov	54
Obrázok 4.2.1 Dimenzie odchodu z organizovaného futbalu	60
Obrázok 4.2.2 Návrh dátových zdrojov a postupu pre zvýšenie dimenzie výstupu	63
Obrázok 5.1.1 Vysvetlenia dát – aktuálne vs navrhované dáta	66

Zoznam grafov

Graf 3.4.1 Priemerný vek registrácie športovcov do SFZ - rok narodenia 1999 - 2009	41
Graf 3.4.2 Priemerný vek žiakov pri registrácii za sezónu podľa závislej premennej ...	42
Graf 3.4.3 Priemerný počet nominácií na zápas za sezónu podľa závislej premennej .	42
Graf 3.4.4 Priemerný počet zápasov v základnej zostave podľa závislej premennej ...	43
Graf 3.4.5 Priemerný počet gólov za sezónu podľa závislej premennej	43
Graf 3.4.6 % podiel žiakov na viacerých pozíciách za sezóny.....	44
Graf 3.4.7 % podiel žiakov na pozícii záložník za sezónu podľa závislej premennej	44
Graf 3.4.8 Zobrazenie rozdielov ostatných nezávislých premenných	45
Graf 3.5.1 Priemerná kvalifikácia trénerov CTM 2015-2021	46
Graf 3.5.2 Kategorizovanie dôvodov odchodu športovcov z CTM 2015-2021	47
Graf 3.5.3 Vplyv počtu trénerov na počet športovcov v CTM	49
Graf 4.2.1 Kategorizovanie dôvodu odchodu športovcov – návrh	62
Graf 5.1.1 Predpokladaná miera kategorizácie odchodu podľa scenárov	65
Graf 5.2.1 Náklady návrhu získania potrebných dát podľa scenárov	67

Zoznam tabuliek

Tabuľka 3.3.1	Finálne čistenie dát od odchýlok v MySQL	40
Tabuľka 3.6.1	Kódovanie závislých a nezávislých premenných	49
Tabuľka 3.6.2	Základné štatistické údaje tréningovej množiny	50
Tabuľka 3.6.3	Základné štatistické údaje testovacej množiny	50
Tabuľka 3.6.4	Výstup logistického modelu - všetky premenné	52
Tabuľka 3.6.5	Výstup logistického modelu - štatisticky významné premenné	52
Tabuľka 3.6.6	Výstup logistického modelu - korelované premenné	53
Tabuľka 3.6.7	Výsledky tuningu SVM modelu - kernel trik radial	55
Tabuľka 3.6.8	Výsledky tuningu SVM modelu - kernel trik polynomial	55
Tabuľka 4.1.1	Sumarizácia scenárov získavania dát	59
Tabuľka 4.2.1	Mapovanie dôvodu odchodu z CTM na návrh	60
Tabuľka 4.2.2	Výsledky pohybu kategórie žiaci v SFZ	61
Tabuľka 4.2.3	Sumarizácia scenárov zvyšovania dimenzií výstupu	64
Tabuľka 5.2.1	Náklady návrhu zvýšenia počtu dimenzií podľa scenárov	68

Zoznam skratiek

Skratka Anglický význam

Bal. Acc	Balance Accuracy
CTM	Center for Talented Youth
DB	Database
ERD	Entity-relationship Model
IoT	Internet Of Things
IT	Information technology
LMŽ	Under-13
LSŽ	Under-15
MINEDU	Ministry of Education, Science, Research and Sports of the Slovak Republic
PII	Personal Identification Information
SFZ	Slovak Football Association
SPI	Sensitive Personal Information
SŠ	High school
SsFZ	Central Slovak Football Association
SŠŠ	High School for Athletes
SVM	Support vector machine
SQL	Structured Query Language
VsFZ	Eastern Slovak Football Association
ZŠ	Elementary school
ZsFZ	West Slovak Football Association

Slovenský význam

Presnosť vyváženia modelu
Centra talentovanej mládeže
Databáza
Entitne-vzťahový diagram
Internet vecí
Informačné technológie
Liga mladších žiakov
Liga starších žiakov
Ministerstvo školstva, vedy, výskumu a športu Slovenskej republiky
Osobné identifikačné údaje
Slovenský futbalový zväz
Citlivé osobné údaje
Stredná škola
Stredoslovenský futbalový zväz
Stredná športová škola
Metóda podporných vektorov
Štruktúrovaný vyhľadávací jazyk
Východoslovenský futbalový zväz
Základná škola
Západoslovenský futbalový zväz

ÚVOD

Šport predstavuje dôležitú súčasť ľudského života. Zohráva zásadnú úlohu počas individuálneho, ale aj spoločenského rozvoja osobnosti. Je spojený s viacerými oblasťami, ktoré udržiavajú ľudí zdravými po fyzickej, ale aj duševnej stránke. Šport môže byť využitý ako nástroj pre zlepšovanie kardiovaskulárneho zdravia populácie, svalovej vytrvalosti a sily alebo zlepšovania flexibility. Tieto oblasti sú extrémne dôležité, keď sa do úvahy zoberie fakt zvyšovania sedavého životného štýlu. Človek je sociálna bytosť, ktorá potrebuje interakciu s inými ľudskými bytosťami. Tímový šport poskytuje skvelý základ pre získanie komunikačných schopností a naučenie jedincov spolupracovať za účelom dosiahnutia spoločného cieľa. Človek je zložený z viacerých častí, ktoré ho robia unikátnym. Počas dospievania je nepochybne jednou z týchto častí aj šport, ktorý pomáha vytvoriť u detí mnohodomenzionálnu identitu, čím im pomáha k vytvoreniu komplexnejšej osobnosti. V neposlednom rade má šport pozitívny vplyv na spoločnosť ako celok. Spája ľudí a podporuje pocit komunity – „patrenia niekam“.

Futbal svojou dostupnosťou a popularitou po celom svete tvorí veľkú časť športu. Na Slovensku futbal predstavuje najobľúbenejší a najpopulárnejší šport medzi všetkými vekovými kategóriami. Z dôvodov vyššie uvedených a iných sa v diplomovej práci bude rozoberať práve problematika v oblasti športového manažmentu a to odchod detí a mládeže z organizovaného futbalu v Slovenskom futbalovom zväze po dovŕšení 15. roku života. Trend odchodu mládeže zo športu je všeobecne známy. Môžu za to rôzne faktory, ktoré sa v práci budú analyzovať. Cieľom práce je na základe poskytnutých dát Slovenským futbalovým zväzom zistiť, prečo mládež odchádza z organizovaného futbalu, a to s využitím modelového riešenia.

V prvej kapitole diplomovej práce budú uvedené základné pojmy a teoretické východiská problematiky od domácich, ale aj zahraničných autorov. Budú rozoberané metódy, postupy a techniky spracovania, analýzy a modelovania dát. Taktiež sa definujú modelové riešenia iných autorov. V kapitole č. 2 bude načrtnutá metodológia práce a spomenú sa metódy, postupy a techniky, ktoré pomôžu dosiahnuť deklarované ciele. Kapitola č. 3 pozostáva zo 6 častí, a to: definície problematiky, faktorov vplývajúcich na odchod z organizovaného športu, spracovania dát – vytvorenia dátového skladu, ťaženie spracovaných dát, analyzovania reportu Šport v číslach a modelovania. Kapitola č. 3 predstavuje nosný pilier celej diplomovej práce. Kapitola č. 4 sa zaoberá vytvorenými návrhmi a odporučeniami, ktoré sú vytvorené na základe kapitoly č. 3 – analytická časť. Pre návrhy sa vypracujú optimistické i pesimistické scenáre. V poslednej kapitole č. 5 sa rozoberá ekonomické a mimoekonomické vyhodnotenie návrhov. V tejto časti sa autor zameria na stanovenie rozsahu nákladov na realizáciu pre definované scenáre.

1 TEORETICKÉ VÝCHODISKÁ

Kapitola *Teoretické východiská* obsahuje základné východiská pre vypracovanie diplomovej práce, metódy a techniky analýzy dát vo všeobecnosti a taktiež špecificky zamerané na využitie v športovom manažmente. Definujú sa dôležité pojmy pre pochopenie a následný ďalší postup v riešení **problematiky odchodu mládeže po 15. veku narodenia**, prechodu z kategórie žiak do kategórie dorasteneec.

1.1 Športový manažment

Športový manažment predstavuje **mnohostrannú a dynamickú oblasť**, ktorá vyžaduje **hlboké pochopenie biznis logiky** tak zložitého odvetvia, akým je šport. Ako každý manažment aj športový manažment je zložený zo základných činností, medzi ktoré patrí **plánovanie, organizovanie, riadenie a kontrolovanie**. Obsahuje širokú škálu funkcií ako je marketing, financie, ľudské zdroje a riadenie viacerých častí organizácie. (PEDERSEN, P. M. a kol. 2017) Jednou z kľúčových výziev športového manažmentu je nájdenie rovnováhy medzi **túžbou po víťazstve a potrebou prevádzkovať životaschopnú organizáciu**. Je preto dôležité správne pochopenie rôznych druhov príjmov vyplývajúcich zo športu, medzi ktoré patrí napríklad aj predaj vstupeniek, sponzorstvo a mediálne práva. (FRIED, G. a kol. 2008) Športový manažment podľa medzinárodného denníka športového manažmentu a marketingu predstavuje „*štúdium a uplatňovanie princípov riadenia na športové organizácie apodniky zapojené do výroby, spotreby a propagácie športu*“. (DORGHAM, M. A. 2021)

Autorsky sa tak športový manažment pre účel práce môže definovať ako **súbor činností** plánovania, organizovania, riadenia a kontrolovania v oblasti športu, kde je obzvlášť dôležité **pochopenie biznis logiky** pre **nájdenie balansu** medzi naplnením túžby vyhrávať zo strany zainteresovaných subjektov a zabezpečenia prevádzkovej schopnosti športovej organizácie.

1.1.1 Dátová analýza v športovom manažmente

Organizácie vo všetkých oblastiach pôsobenia prechádzajú digitálnou transformáciou. Rovnako aj **športový sektor**, a to **na všetkých úrovniach** od amatérskych až po elitné športy (najvyššia úroveň). Hlavnými **zainteresovanými osobami** v sektore športu sú športoví manažéri, obchodní partneri, fanúšikovia, zákazníci, zamestnanci a ďalší. Šport sa transformoval množstvom nových technológií, medzi ktoré patria: **vykonávanie analýzy v pamäti, vizualizácia dát v reálnom čase**, zavedenie **cloud technológie** (zrýchlenie dopytovania), **kamerové záznamy** alebo **senzory** (implementácia IoT). Vymenované technológie umožnili športovcom, klubom, organizáciám a zväzom **ľahšie a efektívnejšie posúdiť výkon športovcov a tímu**. Športovci vďaka prehľadu o svojom vlastnom výkone a výkone ostatných športovcov môžu **trénovať inteligentnejšie, efektívnejšie a špecializovanejšie**. Tréneri dostávajú **informácie vo väčšom množstve a lepšej kvalite**, čo umožňuje vytvorenie optimálnej stratégie pre dosiahnutie želaného výsledku. (STROBEL, T. a kol. 2021)

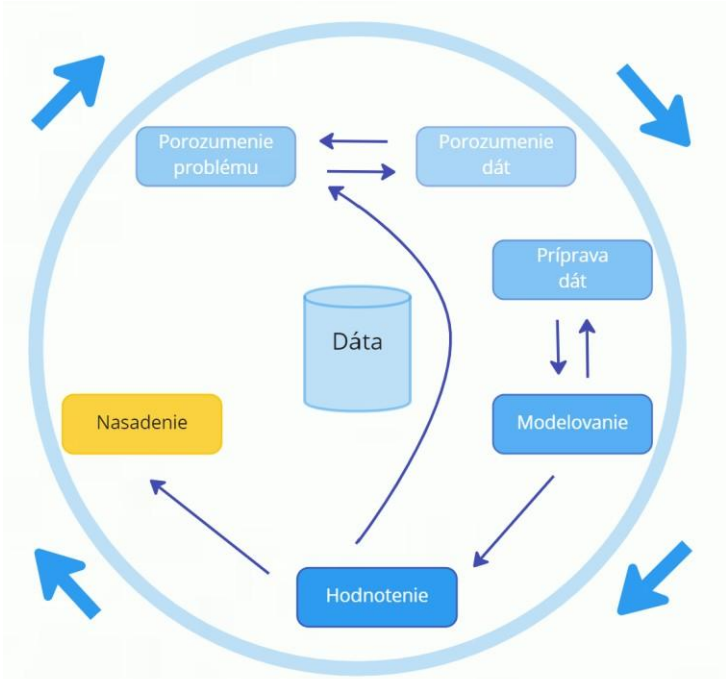
Analýza dát je základnou súčasťou športového manažmentu, pretože umožňuje tímom a organizáciám prijímať informované rozhodnutia na základe vedomostí, znalostí a informácií z údajov. Športová analytika zahŕňa zhromaždenie, spracovanie a analýzu údajov týkajúcich sa rôznych aspektov športu. Medzi tieto aspekty patrí napríklad **výkon hráčov, zapojenie fanúšikov** alebo aj **marketingová stratégia**. V aktuálnom období je využívanie dátovej analýzy v športe primárne zamerané na „big data“ (veľké, rýchle, a neštruktúrované údaje), ale stále sa získavajú informácie aj zo **štruktúrovaných dát**. (SCHUMAKER, P. R. a kol. 2010) Jednou z najdôležitejších techník analýzy údajov v športe je analýza výkonu. **Analýza výkonnosti hráča** je proces zhromažďovania a analýzy údajov o výkonnosti hráča s **cieľom identifikovať vzory, trendy**

a **príležitosti** na zlepšenie. Výkonnostnú analýzu následne tréneri môžu využiť na hodnotenie výkonu jednotlivých hráčov, identifikáciu oblastí na zlepšenie, ale aj ako podklad na tvorbu optimálnej stratégie. Pomocou analýzy výkonnosti môžu tréneri, kluby alebo organizácie robiť informované rozhodnutia o výbere tímu, striedaní hráčov alebo výbere stratégie. (ZHONGBO, B. a kol. 2021)

Ďalšou dôležitou oblasťou analýzy je **angažovanosť fanúšikov**. Fanúšikovská základňa sa podieľa na veľkej časti príjmu klubu, organizácie. Z toho dôvodu je dôležité, aby organizácia mala informácie o správaní sa fanúšikov s cieľom pochopiť ich preferencie, záujmy a motivácie. Výsledky analýzy organizácia môže následne použiť ako podklad pre rozhodovanie v oblastiach efektívnej marketingovej stratégie, zlepšenia zážitku fanúšikov alebo zvýšenia predaja vstupeniek. Športové organizácie tiež používajú analýzu údajov na vyhodnotenie **sponzorstiev** a **partnerstiev**, kde počas analýzy organizácia zhromažďuje dáta o hodnote sponzorstva a účinnosti partnerstiev. Ďalšími metrikami, ktoré sú využívané počas analýzy športových dát sú: počet zobrazení na sociálnych médiách, hodnota značky alebo zapojenie fanúšikov. Analýzou údajov o sponzorstvách, športová organizácia identifikuje, ktoré sponzorstvá a partnerstvá sú kľúčové pre organizáciu a jej zvyšovanie príjmov a môže vykonávať **dátovo založené rozhodnutia**. (SCHUMAKER, P. R. a kol. 2010)

1.2 Ťaženie dát

Ťažba informácií a znalostí z databáz je mnohými výskumníkmi uznaná ako kľúčová výskumná téma v **databázových systémoch** a **strojovom učení**. Spoločnosti a organizácie označili oblasť *ťaženia dát* ako dôležitú, s **príležitosťou na zvýšenie príjmov**. Objavené poznatky z databázy možno aplikovať na správu informácií, spracovanie dotazov, **rozhodovanie**, **riadenie procesov** a mnohé ďalšie využitia. (NGUYEN, G. a kol. 2019) Ťaženie dát, ktoré sa v databázach označuje aj ako **objavovanie znalostí**, predstavuje proces netriviálnej extrakcie **implicitných**, predtým **neznámych** a **potenciálne užitočných informácií** (medzi ktoré patria napríklad pravidlá znalostí, obmedzenia, zákonitosti, atď.) z dát v databázach. (CHEN, M. S. a kol. 2015) **Dopyt po databázových systémoch neustále rastie**. Hlavnými dôvodmi sú **cenová dostupnosť**, **výkonnosť** a **využitelnosť**. Prudký nárast dát a databáz vyvolal **potrebu nových techník** a **nástrojov**, ktoré dokážu spracovať **efektívne množstvo údajov na užitočné informácie**. V dôsledku vývoja situácie na trhu sa ťaženie dát stalo predmetom výskumu a oblasťou s rastúcim záujmom. (PADHY, N. a kol., 2012)



Obrázok 1.2.1 Cyklus ľaženia dát CRISP-DM

Zdroj: (AZEVEDO A. a kol., 2008, s. 184)

Na Obrázok 1.2.1 sa nachádza *cyklus ťaženia dát*, ktorý pozostáva zo **šiestich hlavných fáz**. Cyklus si vyžaduje **iteratívny prechod**, čo znamená pohyblivosť medzi jednotlivými fázami, oboma smermi (dopredu - dozadu). Cyklus sa označuje aj ako *koncept CRISP-DM*. (FERNANDEZ, G. 2011) Výsledok cyklu *závisí od kvality každej jednej fázy*, pričom hlavnými fázami sú:

- **Obchodné porozumenie** – fáza zameraná na pochopenie obchodných cieľov, logiky a požiadaviek z hľadiska riešenia problematiky. Zistené obchodné znalosti slúžia ako **základ pre správne definovanie problematiky** ľaženia dát. Výsledkom etapy obchodné porozumenie je **návrh predbežného plánu** na dosiahnutie cieľov.
- **Pochopenie dát** – fáza pozostáva primárne z **definície zdrojov**, oboznámenia sa so **štruktúrou a vzťahmi medzi dátami** a zdrojmi, vytvorenia prvých pohľadov na dáta, **vytvorenia podmnožín a stanovenia hypotéz**.
- **Príprava dát** – fáza je charakteristická **ETL procesom**, čiže extrakciou dát definovaných zdrojov. **Transformáciou dát** prostredníctvom **úpravy** (pracovanie s chýbajúcimi hodnotami, citlivými údajmi, typom dát, detrendovanie, atď.) a **odstraňovania** (znížením počtu dát nepotrebných dát pre definovaný problém, atď.).
- **Modelovanie** – **výber a aplikácia metód a techník** modelovania, **kalibrovanie parametrov** za účelom dosiahnutia **optimálneho modelu**, ktorý vysvetľuje definovaný problém.
- **Hodnotenie** – po vytvorení sa model **vyhodnotí** a prehodnotí sa jeho aplikácia. Výsledky modelu sa porovnajú s definovanými kľúčovými ukazovateľmi a určí sa **miera splnenia definovaných cieľov** z prvej fázy cyklu. Model môže vyriešiť komplementárne problémy, poskytnúť odpovede na iné otázky, určiť správne otázky alebo **odhaliť neriešiteľnosť stanoveného problému s dostupnými dátami**.
- **Nasadenie** – hlavným cieľom modelu je **zvýšiť vedomosť** a získať nové relevantné poznatky. Výsledky potrebujú byť prezentované v pochopiteľnej podobe aj pre neodborníkov, aby sa zistenia a vedomosti mohli aplikovať v biznise. Výsledky môžu byť prezentované rôznymi formami napríklad ako **jednoducho generovaná správa**

(RMarkdown) alebo prostredníctvom zložitej implementácie **opakovaného procesu** ťaženia dát v rámci celej organizácie (PowerBI, Tableau – informačné prehľady). (NGUYEN, G. a kol. 2019), (PADHY, N. a kol. 2012)

Každá fáza *ťaženia dát* predstavuje dôležitú súčasť, bez ktorej nie je možné dosiahnutie stanoveného cieľu. Najväčšie a najčastejšie problémy vznikajú v dôsledku **zlej definície cieľov** a problematiky, ktorá má byť *procesom ťaženia dát* vysvetlená. Príčinou môže byť rozdiel medzi definíciou cieľov manažmentom a vykonávateľom procesu, „baníkom dát“. (PANG, B. a kol., 2008) Z toho dôvodu je obzvlášť dôležitá fáza pochopenia problému skúmanej problematiky, aby sa následne mohli nájsť správne riešenia problému, modelovania a interpretácie výsledkov.

Ťaženie dát môže byť **rozdelené podľa vykonávaných úloh** v procese ťaženia, ktoré sú zvolené v závislosti od **využitia výstupu**. Jednotlivé úlohy sa klasifikujú na:

- **Prieskumná analýza dát** – úložiská dát (databázy) obsahujú obrovské množstvo údajov. Prieskumnej analýzy dát je založená na **hľadaní vedomostí v dátach bez konkrétneho zámeru**. Z toho dôvodu je prieskumná analýza primárne **vizuálna** a **interaktívna**. Najčastejšie sa využíva vo fáze pochopenia dát, kde nie sú ešte jasné vzťahy a pravidlá medzi dátami.
- **Opisné modely** – popisujú dáta pomocou **rozdelenia p-rozmerného priestoru do skupín** alebo vytvorením **modelov popisujúcich vzťahy medzi premennými**.
- **Prediktívne modelovanie** – **regresná** alebo **klasifikačná** analýza, prostredníctvom ktorej jedna alebo viaceré nezávislé premenné predpovedajú závislú premennú.
- **Objavovanie vzorov a pravidiel** – objavenie **skrytého vzoru** alebo **zhlukov dát**. Dosahuje sa napríklad prostredníctvom **indukcie pravidiel** alebo iných techník **využitia algoritmov** v procese ťaženia dát, ako je aj napríklad „K-Means“.
- **Načítanie podľa obsahu** – cieľom načítania podľa obsahu je nájsť definovaný súbor dát. Často sa používa na vyhľadanie dát v obrázkoch, audio alebo video obsahu. Zameriava sa na hľadanie vzorov, ktoré sú podobné vzoru záujmu. (LAROSE, T. D. a kol. 2014), (NGUYEN, G. a kol. 2019)

1.2.1 Dolovanie vedomostí z databázy

Ťaženie dát je závislé od aplikácie, kde rôzne aplikácie si vyžadujú rôzne **techniky ťažby**. Vo všeobecnosti sa druhy vedomostí, ktoré sú objavené v databáze dajú kategorizovať nasledovne:

Ťažobné **asociačné pravidlá** v transakčných alebo relačných databázach pritiahli veľkú pozornosť v databázových komunitách. Napríklad z veľkého súboru údajov o transakciách možno nájsť spojenia alebo pravidlá tak, že ak si zákazník kúpi (jednu značku) mlieka, zvyčajne si kúpi chlieb (inú značku) v rámci tej istej transakcie. Keďže pravidlá asociácie ťažby dát môžu vyžadovať opakované skenovanie cez veľkú databázu transakcií s cieľom nájsť rôzne **vzory** a **asociácie**, najväčším problémom je **spracovanie dát**, kde základným cieľom je zlepšenie výkonu. (CHEN, M.S. a kol. 2015)

Najpopulárnejšími nástrojmi na ťaženie a analýzu dát, ktoré sú spojené s databázovými systémami, sú nástroje na **zovšeobecňovanie** a **sumarizáciu dát**. Tieto nástroje nesú niekoľko alternatívnych názvov, ale najpoužívanejšími sú **on-line analytické spracovanie** (OLAP), **zovšeobecňovanie**, **sumarizácia**, **abstrakcia dát**, **viacrozmerné databázy** a pod.. Zovšeobecňovanie a sumarizácia poskytujú všeobecné charakteristiky alebo **súhrnný pohľad** na vysokej úrovni, na súbor špecifikovaných dát v databáze. (CHEN, M.S. a kol. 2015)

Ďalšou dôležitou aplikáciou ťaženia dát je schopnosť vykonávať **klasifikáciu** v obrovskom množstve dát. Výsledkom procesu sú **pravidlá klasifikácie ťažby**.

Klasifikácia dát predstavuje rozdelenie súboru údajov na základe ich hodnôt podľa určitých atribútov. Napríklad je požadované, aby predajca áut klasifikoval svojich zákazníkov podľa preferencií pre autá, aby obchodníci vedeli, na koho sa majú obrátiť a katalógy nových modelov sa mohli posilať poštou personalizovane zákazníkom s identifikovanými vlastnosťami, aby sa maximalizovala obchodná príležitosť. (CHEN, M.S. a kol. 2015), (PANG, B. a kol. 2008)

Klastrovanie údajov v podstate znamená zoskupenie množiny údajov (bez preddefinovaného atribútu – učenie bez učiteľa) na základe princípu **konceptuálneho zhľukovania**: maximalizácia podobnosti vnútri triedy a minimalizácia podobnosti z n-tej triedy. Napríklad množinu komoditných objektov je možné najskôr zoskupiť do množiny tried a potom možno na základe takejto klasifikácie odvodiť **súbor pravidiel**. Takéto zhľukovanie môže uľahčiť tvorbu **taxonómie**, čo predstavuje organizovanie pozorovaní do hierarchie tried, ktoré sú zoskupené podľa podobných udalostí. (CHEN, M.S. a kol. 2015), (PANG, B. a kol., 2008)

Časové alebo priestorovo-časové dát tvoria veľkú časť údajov uložených v počítačoch. Príklady tohto typu databázy zahŕňajú: finančnú databázu pre index cien akcií, lekárske a multimediálne databázy, hľadanie podobných vzorov v časovej alebo priestorovo-časovej databáze. Operácie v tomto type dát sú založené na: **objavení a predpovedaní rizika, kauzality alebo trendu** spojeného s konkrétnym vzorom. Typické otázky zahŕňajú identifikáciu spoločností s podobnými modelmi rastu, produktov s podobnými modelmi predaja, akcií s podobným pohybom cien na trhu, obrázkov počasias totožnými poveternostnými podmienkami, geologickými charakteristikami, znečistením prostredia, astrofyzikálnymi modelmi a pod.. Prakticky cieľom pri tomto type dát je vysvetliť, ktoré **nezávislé premenné** (príčina) vplyvajú na **závislú premennú** (následok). (CHEN, M.S. a kol. 2015), (PANG, B. a kol. 2008)

Okrem vyššie uvedených kategórií riešenia problémov existujú aj ďalšie metódy a techniky ťaženia dát, ktoré sa oplatí využívať v určitých prípadoch. Často je potrebné počas procesu ťaženia dát **použiť štruktúrovaný dopytovací jazyk** (SQL) na ťaženie dát alebo **grafické používateľské rozhranie** na špecifikovanie zaujímavej podmnožiny údajov, príslušnej sady atribútov a druhov pravidiel, ktoré sa majú objaviť. Okrem toho je často potrebné vykonávať interaktívne ťaženie dát na preskúmanie, transformáciu a manipuláciu s priebežnými výsledkami ťaženia dát, zamerať sa na rôzne úrovne **konceptov** alebo testovať rôzne druhy **prahových hodnôt** (najnižšia hodnota, ktorá spadá ešte do vytvorenej skupiny/kategórie, určená podľa zoradenia dát). **Vizuálna reprezentácia dát a znalostí** (PowerBI, Tableau) môže uľahčiť *interaktívne získavanie znalostí* v databázach. (CHEN, M.S. a kol. 2015), (PANG, B. a kol. 2008)

1.2.2 Využitie ťaženia dát pre podporu rozhodovania

Získavanie informácií, vedomostí a znalostí z údajov predstavuje aktuálne najdôležitejší zdroj podniku alebo organizácie, pričom vo výraznej miere prispieva pri k získaní **konkurenčnej výhody** alebo **potrebných odpovedí**, za predpokladu správnej aplikácie. (CHEN, M.S. a kol. 2015) **Dáta rastú exponenciálne** a techniky, metódy a postupy pre ich spracovanie sa vyvíjajú rýchlym tempom. (STATISTA [online] 2023) Každý podnik má v určitej *forme databázu* (dátový sklad) alebo *databázy*, kde ukladá všetky dáta, ktoré zhromaždil počas svojej činnosti. Aplikovaním ťaženia dát v podniku pri **interných procesoch** (výrobe, komunikácii, riadení, atď.) alebo **externých procesoch** (marketing, vzťahy s verejnosťou, vývojom trhu, atď.) poskytuje za predpokladu, že všetky fázy procesu ťaženia dát sú vykonané správne, **najlepšie podklady pre výber optimálneho rozhodnutia založeného na dátach**. (NGUYEN, G. a kol. 2019)

Ťaženie dát poskytuje produkty od rôznych spoločností, ktorých funkcionality sa môže rozdeliť, na základe **typu používateľa**. Technické typy používateľov, zvyčajne

dátoví vedci, dátoví analytici alebo databázoví špecialisti využívajú pokročilé nástroje ťaženia dát, kde vytvárajú **reporty** alebo **prehľady, grafy, interaktívne informačné panely, tabuľky**. (LAROSE, T.D. a kol. 2014) Výstup technického používateľa využívajú manažéri pre úpravu stratégie alebo výber optimálneho riešenia problému. Netechnické typy, zvyčajne manažéri na všetkých úrovniach, využívajú *nástroje*, ktoré nevyžadujú žiadne alebo len základné programátorské znalosti a poskytujú parciálne **informácie potrebné pre rozhodovanie**. (FERNANDEZ G. 2011)

V každej *oblasti* alebo *divízii podniku* je potrebné robiť rozhodnutia. **Rozhodnutia založené na dátach** predstavujú **optimálny variant pre dosiahnutie stanovených cieľov**. Je však potrebné zdôrazniť, že získavanie znalostí z databáz alebo ťaženie dát, je len tak správne, ako sú správne dáta, na ktorých sa proces vykonáva. Veľké množstvo podnikov a organizácií, ktoré prechádza *na ťaženie dát*, má **problém s procesom získavania, transformácie a načítania dát (ETL)**, ktorý je nevyhnutnou podmienkou pred začatím procesu *ťaženia dát*. (NGUYEN, G. a kol. 2019)

1.3 Dátové modelovanie

Dátové modelovanie predstavuje silný nástroj na **predpovedanie** budúceho správania/stavu **na základe údajov z minulosti**. Vytvorené modely možno použiť najmä na relatívne presné predpovede budúcich výsledkov, ktoré sú nesmierne dôležité v rôznych oblastiach od financií cez ekonomiku až po zdravotníctvo a inžinierstvo. Predikčné modely môže využiť **široké množstvo štatistických nástrojov, techník a metód strojového učenia**. Predtým ako sa vytvorí model je dôležité najskôr identifikovať problém, ktorého riešením má byť model. Následne je potrebné zistiť aké dátové zdroje sú dostupné na riešenie problému. Až po splnení týchto základných faktorov je možné zvoliť vhodné metódy a techniky modelovania. (HYNDMAN, R. J. a kol. 2021)

Pri výbere metód a techník modelovania je dôležité definovať, aký **typ modelu učenia** sa využije. Kritériom pre určenie typu modelu učenia je existencia, respektíve neexistencia výstupu, ktorý slúži ako kontrola výkonnosti modelu. Modely sa následne z toho hľadiska dajú rozdeliť na: **učenie s dohľadom** (supervised learning) alebo **učenie bez dohľadu** (unsupervised learning). Učenie s dohľadom prebieha formou **trénovania funkcie** na predpovedanie výstupu na **základe vstupných dát**. Výstup môže byť buď **klasifikácia/triedy** (klasifikačný problém) alebo **predpovedanie výsledku** (regresný algoritmus). Učenie bez dozoru **odhaľuje základný vzor** v súbore údajov, ktorý **nie je explicitne prezentovaný**. Môže tak odhaliť podobnosť údajových bodov (algoritmy klastrovania) alebo skryté vzťahy premenných (algoritmy asociačných pravidiel). **Posilňovacie učenie** predstavuje samostatný typ strojového učenia, kde **agenti sa učia konať na základe interakcie s prostredím s cieľom maximalizovať odmeny**. Tento spôsob učenia sa podobá procesu učenia človeka, ktorý sa riadi metódou pokus-omyl. (KUHN, M. a kol. 2013)

1.3.1 Učenie s dozorom

Učenie s dozorom (supervised learning) je typ strojového učenia, kde algoritmus je trénovaný na údajoch, ktoré sú označené, aby mohol **predpovedať** alebo **rozhodovať o nových** (testovacích) **údajoch**, ktoré nie sú označené. Cieľom učenia s dozorom je naučiť sa **mapovanie zo vstupov na výstupy** na základe súboru dát, kde pri trénovaní každý príklad zo súboru dát pozostáva z vstupov a zodpovedajúceho výstupu. Algoritmus potom použije vytvorené mapovanie na predpovedanie alebo rozhodnutie o nových údajoch, ktoré už poskytujú len vstupy. (KUHN, M. a kol. 2013) Učenie s dozorom sa využíva v mnohých oblastiach, medzi ktoré patria napríklad aj financie alebo zdravotníctvo. Ako príklady použitia možno uviesť: predpovedanie rozhodnutia zákazníka o kúpe produktu na základe jeho nákupného správania v minulosti, diagnostika chorôb na základe lekárskeho obrázkov alebo klasifikácia

textových dokumentov na základe obsahu. (GOODFELLOW, I. a kol. 2016) Učenie s dohľadom je ďalej možné kategorizovať na **klasifikačné** a **regresné algoritmy**. Klasifikačné modely identifikujú, do ktorej kategórie objekt patrí. Na druhej strane regresný model **predpovedá nepretržitý výstup**. Klasifikácia využíva regresný model s aplikovanou **prahovou hodnotou** (threshold value). Prahová hodnota sa používa na **klasifikátor**, kde v prípade, ak je výstup modelu vyšší ako prahová hodnota, objekt je klasifikovaný ako pravdivý (1). V prípade nižšieho výstupu je objekt klasifikovaný ako nepravdivý (0). (KUHN, M. a kol. 2013)

Existuje veľké množstvo rôznych algoritmov učenia s dozorom, z ktorých každý má silné stránky, ale aj limitácie. Medzi **najznámejšie algoritmy** učenia s dozorom patria: lineárna regresia, logistická regresia, rozhodovacie stromy, náhodné lesy alebo metóda podporných vektorov (support vector machine). Spomenuté algoritmy možno použiť pre širokú škálu úloh vrátane **klasifikácie, regresie a prognózovania**. (KUHN, M. a kol. 2013)

Na tréning algoritmu učenia s dozorom je potrebný súbor údajov, ktorý obsahuje objekty s vstupmi a označenými výstupmi. Súbor dát je dôležitý pred učením rozdeliť na tréningovú a testovaciu množinu. **Tréningová množina** sa používa na učenie algoritmu, kde výsledkom je vytvorený model, ktorý sa snaží minimalizovať chybovosť predikcií. **Testovacia množina** je využitá na vyhodnotenie výkonu modelu na nových dátach, ktoré vytvorený model ešte nevidel. Rozdelenie súboru dát na tréningový a testovací dataset má hlavný význam z hľadiska zamedzenia **nadmerného prispôsobovania modelu** tréningovým dátam (overfitting). (GOODFELLOW, I. a kol. 2016)

1.3.2 Klasifikačné modely

Klasifikácia predstavuje techniku, pri ktorej sa údaje kategorizujú do určitého počtu kategórií/tried. Úlohou klasifikačného modelu je vytvoriť záver na základe vstupných dát. Záver sa vykonáva na základe vytvoreného **mapovania** počas tréningovania modelu. Výstup na nových dátach predstavuje predpoveď kategórie/triedy. Pri **binárnej klasifikácii** model pracuje s dvoma možnými kategóriami/triedami, a to 0 alebo 1. (HYNDMAN, R. J. a kol. 2021) Existuje viacero typov klasifikačných modelov, pričom tie, s ktorými sa pracuje v tejto diplomovej práci, sú definované nižšie.

Logistická regresia predstavuje algoritmus strojového učenia používaného pre úlohy **binárnej klasifikácie**, kde cieľom je predpovedanie, do ktorej z dvoch tried vstup patrí. Na rozdiel od lineárnej regresie, ktorá predpovedá spojité hodnoty, logistická regresia predpovedá **pravdepodobnosť**, s akou vstup patrí do skupiny. Logistická regresia pripája k tréningovým dátam logistickú funkciu. **Logistická funkcia** predstavuje krivku v tvare S, ktorá mapuje vstupnú hodnotu medzi hodnotou 0 až 1, čo sa interpretuje ako pravdepodobnosť príslušnosti k pozitívnej triede (1). (HASTIE, T. a kol., 2008) Logistická funkcia má tvar:

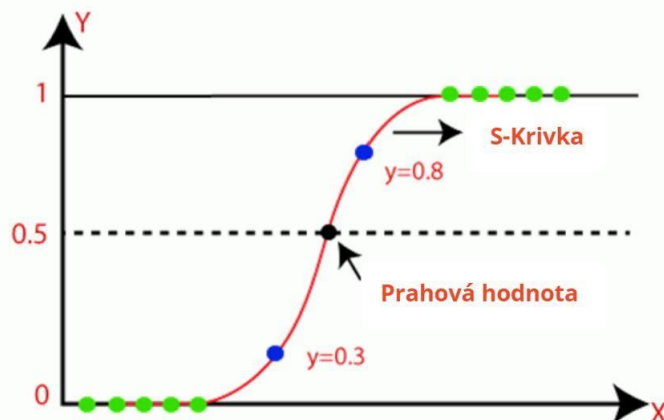
$$p(x) = 1 / (1 + \exp(-z)),$$

kde **p(x)** je *predpokladaná pravdepodobnosť výstupu*, **exp** je *exponenciálna funkcia* a **z** je lineárna kombinácia vstupných znakov, ktorá sa dá zapísať ako:

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n,$$

kde **b₀** je člen priesečníka a **b₁** až **b_n** sú koeficienty vstupných znakov.

Medzi najväčšie výhody *logistickej regresie* patrí **pochopenie vplyvu nezávislých premenných** na závislú premennú. Na druhej strane najväčšou limitáciou je predpoklad, že všetky **vysvetľujúce premenné sú navzájom nezávislé** (žiadna korelácia medzi nezávislými premennými, čiže každá z nich vysvetľuje závislú premennú inak) a **vstupné údaje neobsahujú chýbajúce hodnoty**. (HASTIE, T. a kol., 2008) Na Obrázok 1.3.1 je ilustrovaná S krivka logistickej regresie:



Obrázok 1.3.1 Logistická regresia
Zdroj: (Trevor Hastie a kol. 2008)

Naivný Bayesov klasifikátor predstavuje jednoduchú, ale za to efektívnu klasifikačnú metódu založenú na **Bayesovej vete**. Veta hovorí, že pravdepodobnosť hypotézy pri niektorých pozorovaných dôkazoch je úmerná súčinu predchádzajúcich pravdepodobností hypotézy a pravdepodobnosti dôkazu danej hypotézy. Označenie naivný vychádza z predpokladu, že všetky vstupné funkcie sú od seba **nezávislé vzhľadom na označenie triedy**. (HASTIE, T. a kol., 2008) Naivný Bayesov klasifikátor má rovnicu v tvare:

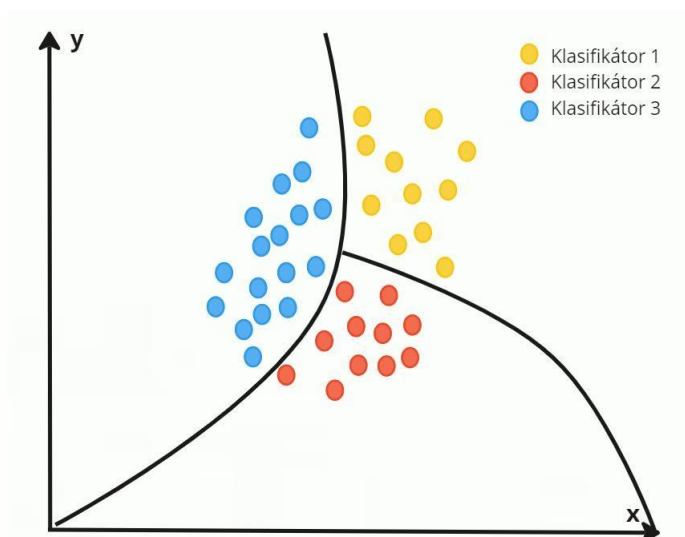
$$P(C|X) = (P(C) * P(X|C)) / P(X),$$

kde **P(C|X)** je zadná pravdepodobnosť triedy C vzhľadom na vstupné znaky X, čo je pravdepodobnosť, ktorá sa počíta pre klasifikáciu. **P(C)** je predchádzajúca pravdepodobnosť triedy C, čo je pravdepodobnosť pozorovania triedy C bez zohľadnenia akýchkoľvek vstupných znakov. **P(X|C)** je pravdepodobnosť vstupných znakov X danej triedy C, čo je pravdepodobnosť pozorovania vstupných znakov X vzhľadom na to, že inštancia patrí do triedy C. **P(X)** je dôkaz alebo hraničná pravdepodobnosť vstupných znakov X, čo je pravdepodobnosť pozorovania vstupných znakov X bez ohľadu na triedu.

$$P(X|C) = P(x_1|C) * P(x_2|C) * \dots * P(x_n|C),$$

kde **x_i** je i-tá vstupná funkcia a **n** je celkový počet vstupných funkcií.

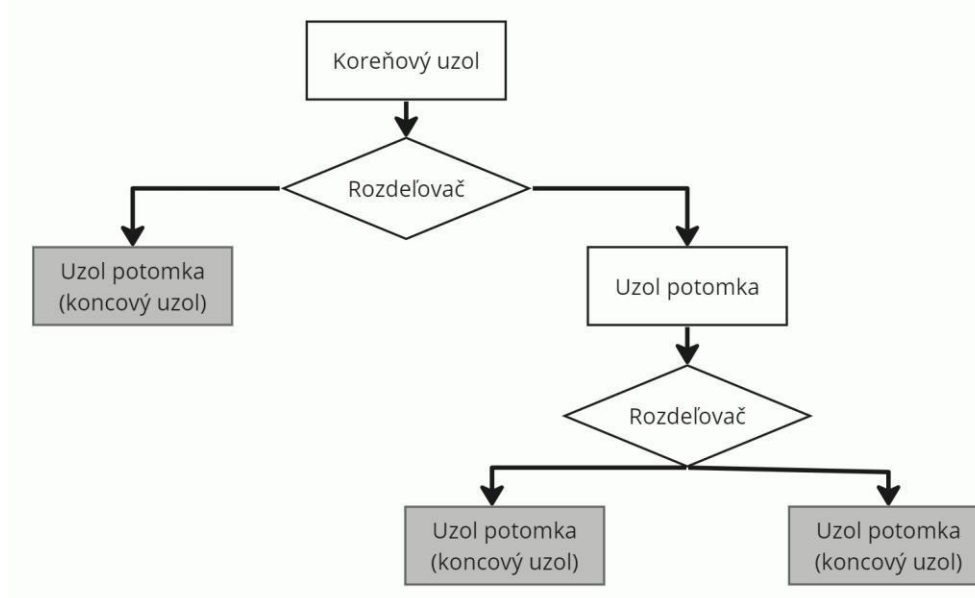
Naivný Bayesov klasifikátor vypočíta **posteriórnu pravdepodobnosť** (revidovaná alebo aktualizovaná pravdepodobnosť udalosti, ktorá nastane po zohľadnení nových informácií) pre každú možnú triedu a vyberie triedu s najväčšou pravdepodobnosťou ako predpovedanú triedu pre dané vstupné znaky X. Predpoklad zjednodušuje výpočet pravdepodobností a robí **metódu výpočtovo efektívnou a škálovateľnou pre vysokorozmerné súbory údajov**. V praxi však predpoklad nezávislosti funkcií nemusí platiť, čo vedie k **neoptimálnemu výkonu** alebo **nesprávnym predpovediam**. (HASTIE, T. a kol., 2008) Na Obrázok 1.3.2 je zobrazený naivný Bayesov klasifikátor.



Obrázok 1.3.2 Naivný Bayesov klasifikátor

Zdroj: *vlastné spracovanie na základe Koushiki Dasgupta Chaudhuri 2022*

Rozhodovacie stromy vytvárajú model na základe údajov o atribútoch spolu s ich kategóriami/triedami, kde algoritmus vytvára **postupnosť pravidiel delenia**, ktoré sú použité na klasifikáciu údajov. Algoritmus funguje na **rekurzívnom rozdelení priestoru funkcie** na menšie a menšie oblasti s cieľom čo najčistejšie oddeliť rôzne triedy. Každý oddiel sa vytvorí výberom funkcie a prahovej hodnoty, ktorá rozdelí údaje do dvoch podmnožín podľa toho, či je hodnota funkcie nad alebo pod prahovou hodnotou. Tento proces prebieha, pokiaľ sa nepríde na koniec stromu. (HYNDMAN, R. J. a kol. 2021) Na **Obrázok 1.3.3** je znázornená ilustrácia rozhodovacieho stromu.

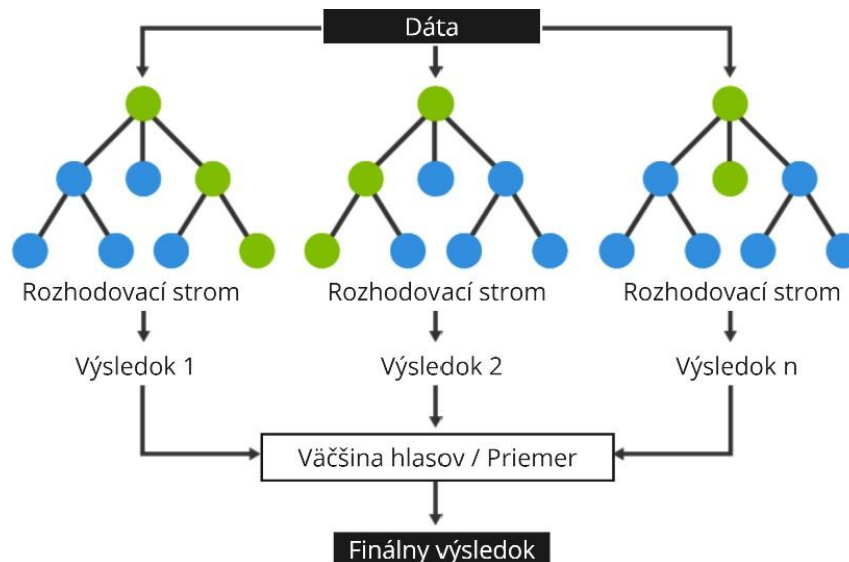


Obrázok 1.3.3 Rozhodovací strom

Zdroj: *vlastné spracovanie na základe Bei Zhou a kol. 2019*

Model je **jednoduchý na vizualizáciu** a **pochopenie pravidiel delenia**, pred vytvorením **nevyžaduje zložitú prípravu** dát a dokáže pracovať s numerickými, rovnako i kategorickými dátami. Medzi najväčšie limitácie patrí však práca s komplexnými dátami, kde sú **vytvorené stromy zložité** a môžu byť **nestabilné**, kde malá odchýlka v dátach vedie k vygenerovaniu úplne iného stromu. To môže viesť k **slabému výkonu** alebo **zovšeobecňovaniu nových údajov**. (HYNDMAN, R. J. a kol. 2021)

Náhodný les predstavuje meta-odhad, ktorý vytvára množstvo rozhodovacích stromov na rôznych parciálnych častiach množín údajov. Počas predikčnej fázy **každý strom v lese** nezávisle predpovedá výstup, kde konečný výstup je určený väčšinovým hlasom (**klasifikácia**) alebo priemerom (**regresia**) jednotlivých stromov. (HYNDMAN, R. J. a kol. 2021)



Obrázok 1.3.4 Náhodný les
Zdroj: *vlastné spracovanie na základe Tibco 2022*

Medzi výhody patrí **nižšia pravdepodobnosť pretrénovania** a model má taktiež vo väčšine prípadov **lepšiu prediktívnu schopnosť** ako *rozhodovacie stromy*. Medzi nevýhody patrí najmä *pomalá predikcia v reálnom čase*, **náročná implementácia** a **zložitý algoritmus**. (HYNDMAN, R. J. a kol. 2021)

Podpora vektorého stroja (support vector machine, ďalej už len „SVM“) predstavuje reprezentáciu tréningových dát ako bodov v priestore, oddelených do kategórií jasnou medzerou (nazývanou aj hyperplane - **nadrovina**), ktorá **maximalizuje rozpätie medzi triedami**. Model je schopný vykonávať ako lineárnu tak aj nelineárnu klasifikáciu. Nové dáta sa následne **mapujú do vytvoreného tréningového priestoru** a predpokladá sa, že patria do kategórie na základe toho, na ktorú stranu medzery spadajú. (HASTIE, T. a kol., 2008)

$$h(x) = \text{sign}(w \cdot x + b),$$

kde **$h(x)$** je predpovedaná trieda nového vstupného dátového bodu x , **w** je váhový vektor, ktorý definuje orientáciu nadroviny (hyperplane), $\cdot$ označuje bodový súčin medzi dvoma vektormi, **b** je výraz odchýlky, ktorý posúva nadrovinu (hyperplane) a **$\text{sign}()$** je funkcia znamienka, ktorá vracia +1 pre kladné hodnoty a -1 pre záporné hodnoty. (HYNDMAN, R. J. a kol. 2021)

Následne na tréningovanie SVM sa musia **nájsť hodnoty w a b** , ktoré **minimalizujú chybu klasifikácie**, a **zároveň maximalizujú rozpätie** medzi týmito dvoma triedami. Nájdenie hodnôt sa dá vyjadriť ako **optimalizačný problém**:

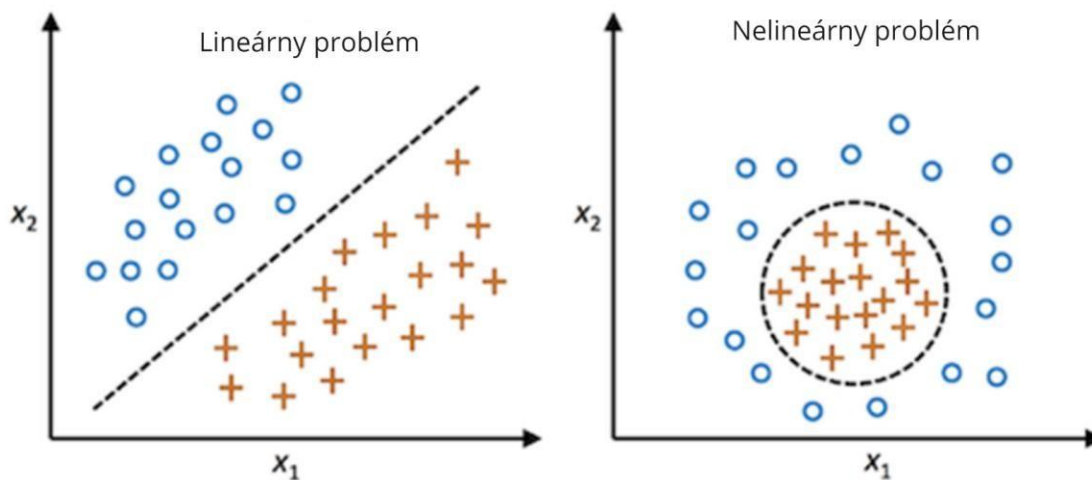
$$\text{minimize } 1/2 ||w||^2$$

$$\text{podmienka: } y_i(w \cdot x_i + b) \geq 1 \text{ pre všetky } i,$$

kde $||w||^2$ je norma L2 váhového vektora, **y_i** je skutočné označenie triedy i -tého príkladu tréningu a **x_i** je vektor vlastností i -tého príkladu tréningu. **Obmedzenie nerovnosti**

zaistuje, že všetky tréningové príklady sú správne klasifikované s okrajom aspoň 1. (HASTIE, T. a kol., 2008)

Pre **nelineárnu SVM** sú **vzorec** a **problém optimalizácie** upravené tak, aby zahŕňali **funkciu jadra** (kernel), ktorá mapuje vstupné údaje do priestoru funkcií vyššej dimenzie. (HASTIE, T. a kol., 2008)



Obrázok 1.3.5 Podpora vektorových strojov: Lineárna a Nelineárna
Zdroj: (Shivam Sharma, 2021)

Na Obrázok 1.3.5 je zobrazený na ilustráciu lineárny a nelineárny problém s vytvorenými triedami. Model je **efektívny vo vysokorozmernom priestore** a využíva podmnožinu tréningových bodov v rozhodovacej funkcii, takže je aj **pamäťovo efektívny**. Obzvlášť efektívny je model SVM pri **nelineárnych riešeniach**. Nevýhodou je, že algoritmus **neposkytuje priamo odhad pravdepodobnosti** ako iné modely. **Odhad pravdepodobnosti** sa počíta pomocou **drahého** (časovo a pamäťovo náročného procesu) päťnásobného krížového overovania. (HASTIE, T. a kol., 2008)

1.4 Využitie modelovania na odchod z organizovaného futbalu

Počas prieskumu dokumentov sa našli **2 štúdie**, ktoré pracujú s odchodom detí a mládeže z organizovaného futbalu. Štruktúra spracovania pozostáva z krátkej definície dát, predstavenia uhla pohľadu autorov na problematiku a následnej zhrnutiavýsledkov štúdie.

Štúdia uskutočnená Van Yperen, Nico W. a kolektívov, publikovaná v roku 2022, sa zamerala na **predikovanie odchodu z organizovaného futbalu v holandských futbalových kluboch**. Vzorka športovcov, na ktorej sa realizoval výskum, predstavuje 1762, kde vekové rozpätie je od 16 do 21 rokov a priemerný vek skúmaných športovcov je na úrovni 17,74. Súčasťou štúdie sú aj ženy (27,1 %).

Hlavným predpokladom realizovaného výskumu sú hypotézy, kde odchod z futbalu predstavuje funkciu hráčovej **radosti z futbalu**, **alternatívnych možností** (vykonávanie iných športov) a **obmedzujúcimi silami** (angažovanosť rodiny v klube a dĺžka investovaného času). Na overenie stanovených hypotéz sa použila **logistická regresia**, kde závislou premennou je odchod z organizovaného futbalu (v krátkodobom horizonte počas 6 mesiacov, v dlhodobom počas 4 rokov, alebo športovci pokračoval s organizovaným futbalom).

Predictor	B	SE B	χ^2	Exp(B) (odds ratio)	95% CI Exp(B)	
					Lower bound	Upper bound
Remainers vs. short-term						
Intercept	0.71	1.85	0.15			
Enjoyment	−0.79**	0.12	42.38**	0.46	0.36	0.58
Alt1: Only football	0.49	0.28	3.19	1.63	0.95	2.80
Alt2: Org. football	−0.55	0.30	3.27	0.58	0.32	1.05
RF1: Family	−0.38	0.26	2.15	0.69	0.42	1.14
RF2: First team	0.19	0.08	5.29	1.21	1.03	1.42
Age	0.05	0.09	0.30	1.05	0.88	1.25
Sex	0.43	0.32	1.76	1.53	0.82	2.87
Remainers vs. long-term						
Intercept	3.46	0.96	13.07			
Enjoyment	−0.31**	0.07	18.93**	0.73	0.64	0.84
Alt1: Only football	0.41*	0.13	9.37*	1.51	1.16	1.96
Alt2: Org. football	−0.39*	0.16	6.21*	0.68	0.50	0.92
RF1: Family	−0.31*	0.12	7.01*	0.73	0.58	0.92
RF2: First team	0.17**	0.04	16.17**	1.18	1.09	1.28
Age	−0.14*	0.05	9.98*	0.87	0.79	0.95
Sex	−0.19	0.14	1.85	0.83	0.63	1.09

Obrázok 1.4.1 Výsledok štúdie Van Yperen a kolektívu 2022

Zdroj: (VAN YPEREN, N. W. a kol. 2022)

Na *Obrázok 1.4.1* sú zobrazené výstupy vytvorených modelov. Z výsledkov vytvoreného celkového logistického modelu (presnosť klasifikácie 72,9 %) vyplýva, že v krátkodobom období **radosť z futbalu** (Enjoyment) predstavuje jedinú štatisticky významnú premennú (na úrovni p-hodnoty < 0,001). Model tiež ukazuje, že v dlhodobom období **alternatívne možnosti** (Alt1 a Alt2) a **obmedzujúce sily** (RF1 a RF2) pomáhajú pri vysvetlení variability predikovaných hodnôt. Finálna interpretácia modelu hovorí, že športovci, ktorí nepokračovali s organizovaným futbalom, **mali menšiu radosť z futbalu, vnímali viac alternatív** v iných športoch a mali **menej rodinných príslušníkov zapojených** do organizovaného futbalu. Športovci, ktorí **začali s futbalom neskôr, majú väčšiu pravdepodobnosť skončenia** s organizovaným futbalom. (VAN YPEREN, N. W. a kol. 2022)

Štúdia Deelen, Ineke a kolektívu vykonala **online výskum** v období 2015 až 2016 medzi dospelými športovcami (13 - 21 rokov), kde celková vzorka športovcov predstavovala 2555. Výskum sa zameriaval na športy **futbal** (1014 pokračujúcich a 346 nepokračujúcich športovcov) a **tenis** (602 pokračujúcich a 593 nepokračujúcich športovcov). Skúmané sú obe pohlavia, kde dataset má približne vyrovnanú distribúciu pohlavia (futbal mal väčšie zastúpenie mužov). Online výskum sa zameriaval na získanie dát ohľadom **charakteristiky respondenta, využívanie času, nedávne zmeny v zamestnaní, domácej alebo školskej situácii**, ale taktiež na **druh motivácie športovcov** (skôr na úlohy alebo ego). Športy sa analyzovali osobitne, pretože sú zistené **výrazné rozdiely medzi športmi**.

Pre jednotlivé časti, ktoré sú popísané vyššie, sú vytvorené individuálne modely (všetky modely predstavujú binárnu logistickú regresiu), kde v prvom modeli sú použité len **sociodemografické faktory**, v druhom modeli sa pridávajú **faktory využitia času**, v treťom modeli sa pridávajú faktory **prostredia** a v poslednom štvrtom modeli sa pridávajú faktory **druhu orientácie na motiváciu**. Prvý model ukázal, že tí, ktorí majú **vyššiu úroveň vzdelania** a tí, ktorí **hrajú vo výberovom tíme**, majú nižšiu pravdepodobnosť skončenia. Druhý model hovorí, že dospelí, ktorí trávajú **viac času dobrovoľnými, spoločenskými aktivitami v klube**, majú menšiu pravdepodobnosť skončenia. Tretí model ukazuje, že **faktory prostredia** významne nevysvetľujú ukončenie hráča s organizovaným futbalom. Štvrtý model ukázal, že oba prípady druhu orientácie na motiváciu (na úlohy alebo ego) znižujú pravdepodobnosť skončenia. (INEKE, D. a kol. 2018)

2 METODOLÓGIA PRÁCE

V časti metodológia práce sa definujú hlavné a čiastkové ciele, metódy, techniky a postupy, ktoré sú použité pri písaní diplomovej práce. V rámci metodológie sa tiež špecifikujú využívané softvérové nástroje a metódy využité pri vytváraní návrhov.

2.1 Cieľ práce

Cieľom diplomovej práce je na základe teoretických východísk v problematike športového manažmentu a dátovej analytiky vypracovať návrh na zlepšenie súčasného stavu v špecifikovanej oblasti, s implementáciou modelového riešenia. Oblasť riešenia sa špecifikovala na analýzu odchodu mládeže z organizovaného futbalu vykonávaného vo Slovenskom futbalovom zväze po dovŕšení 15. roku, prechodu z kategórie žiakov do dorastu. Analýza faktorov vplyvujúcich na odchod poskytuje východisko pre faktory, ktoré vysvetľujú nepokračovanie s organizovaným futbalom v SFZ. Cieľom ľaženia dát je zistiť, ktoré dostupné faktory vysvetľujú odchod z futbalu najlepšie a následne ich použiť pri modelovom riešení problematiky. Následne je možné špecifikovať oblasť zamerania diplomovej práce:

Na základe poskytnutých dát Slovenským futbalovým zväzom zistiť pomocou analýzy dát, prečo mládež odchádza z organizovaného futbalu v SFZ s využitím modelového riešenia problematiky.

Parciálne ciele práce môžu byť definované nasledovne:

- Poskytnúť teoretické východiská a popísať súčasný stav skúmanej problematiky so zohľadnením dátovej analytiky.
- Analyzovať literatúru a odborné články v oblasti problematiky odchodu mládeže a detí z organizovaného športu.
- Analyzovať primárne a sekundárne dátové zdroje a ich väzby medzi sebou.
- Vybrať modelovacie prostriedky vhodné pre riešenie skúmanej problematiky.
- Navrhnuť riešenie pre zlepšenie súčasného stavu v problematike odchodu mládeže a detí z organizovaného športu s využitím modelového prístupu.
- Využiť Open Sourceové programovacie prostriedky pre vytvorenie riešenia.
- Definovať a vyhodnotiť ekonomické a mimoekonomické prínosy navrhovaného riešenia.

Parciálne ciele diplomovej práce vyžadujú na svoje naplnenie podporné ciele, medzi ktoré patrí primárne spracovanie poskytnutých dát Slovenským futbalovým zväzom a získaných dát do štruktúry, ktorá umožňuje ľahký, čo najpresnejší prístup k analýze a modelovaniu skúmanej problematiky.

2.2 Metódy, techniky a postupy

Diplomová práca je zložená z piatich hlavných kapitol. Prvú kapitolu predstavujú teoretické východiská, kde sú spracované a analyzované vedecké články a literatúra od domácich a zahraničných autorov, pôsobiach v skúmanej problematike alebo v oblasti implementovaných metód a techník spracovania, analýzy a modelovania dát. Kapitola tri je najobširnejšia a zaoberá sa spracovaním dát, analýzou faktorov vplyvujúcich na problematiku, ľažením dát z vytvoreného dátového skladu a modelovaním predikčného modelu. Medzi dátové zdroje patrí záloha SFZ, sportnet API a reporty Šport v číslach. Počas ekonomického a mimoekonomického vyhodnotenia sa používa analýza zdrojov dostupných na internete.

Využité softvérové nástroje

Počas tvorby diplomovej práce sú využívané primárne nasledovné softvérové nástroje. Od spoločnosti Microsoft produkty Word, ktorý je primárne využívaný pri analýze odborných článkov a literatúry na zaznamenávanie zistení a informácií. Excel, ktorý sa používa na vykonanie základných analýz a štatistík, spracovanie časti dátových zdrojov, vytvorenie grafických výstupov z analýzy a ľaženia dát. PowerPoint sa primárne využíva na vytvorenie podkladov na prezentovanie zistení a návrhov. MongoDBa MongoDB Compass sa využívajú na spracovanie JSON súborov zálohy SFZa sportnet API dát. Postman sa využíva na posielanie požiadaviek na API server sportnet. Python je využitý na úpravu dát do požadovaného tvaru pred nahratím do MongoDB. RStudio slúži na analýzu, vytvorenie modelov a ich optimalizáciu. Na načítanie, spracovanie, úpravu, transformáciu a export dát sa využíva MySQL.

Metódy využité v diplomovej práci

Návrhy a odporúčania sú tvorené pomocou rôznych metód, ktoré sú popísané nižšie. Základnou metódou, ktorá sa používa naprieč všetkými časťami práce je analýza dát a odbornej literatúry. Následne sa využila syntéza informácií, metódy dedukcie, ale aj indukcie a vo veľkej miere kreatívne myslenie. Pre definovanie problému sa využíva metóda čiernej skrinky, kde sa počas brainstormingu špecifikovali dostupné vstupya želané výstupy. Pri navrhovaní klasifikácie výstupu a zdrojov dát pre návrh sa hlavne využíva syntéza a dedukcia z nadobudnutých znalostí počas analýzy dátových zdrojova odbornej literatúry.

3 ANALYTICKÁ ČASŤ

V časti diplomovej práce Analytická časť sa vykonávajú všetky potrebné druhy analýzy pre vyriešenie problému odchodu detí a mládeže z organizovaného futbalu po dovŕšení 15. roku života. Celá štruktúra procesu analyzovania vychádza z krokov CRISP-DM modelu. Na začiatok sa v podkapitole *Definovanie problému* poskytuje uhol pohľadu na problematiku a odôvodňuje sa podstata problematiky. Následne sa vykoná analýza literatúry a odborných článkov pre zistenie *Faktory vplývajúce na odchod z organizovaného športu* u mládeže. So získanými východiskami sa *Spracovanie dát* prejde sa k *Dobovaniu dát poskytnutých SFZ* a *Analýza dát z reportov „Šport v číslach“*. V poslednej časti sa prejde k *Modelovaniu*. Každá podkapitola analytickej časti obsahuje na konci zhrnutie.

3.1 Definovanie problému

Hlavný problém, ktorým sa v diplomovej práci zaoberáme, predstavuje **odchod detí a mládeže z organizovaného futbalu po dovŕšení 15. roku**. Vek je stanovený na základe obdobia, kedy dieťa prechádza v slovenskom futbale z **kategórie žiak** (6 -14) do **kategórie dorastenec** (14 -18). Problematika sa rozoberá na základe poskytnutých informácií a dát, ktorých spracovanie a obsah je vysvetlený v časti *Spracovanie dát*. V kontexte diplomovej práce sa **odchodom z organizovaného futbalu rozumie odchod zo súťaží organizovaných Slovenským futbalovým zväzom**.

3.1.1 Odchod z organizovaného športu po dovŕšení 15. roku

Organizácia ICOACHKIDS v roku 2021 zverejnila výsledky výskumu, kde predmetom je **analyzovanie 36 rôznych športov v podmienkach rôznych krajín**. Výskum je primárne zameraný na **šport u detí a mládeže**. Výsledky výskumu ukazujú, že u detí a mládeže sa **zvyšuje účasť vo vekových kategóriách U8-U14**, v prípade väčšiny športov. Na druhej strane, vo vekových kategóriách **U14-U18 dochádza k výraznému poklesu** účasti detí a mládeže na športe. V prípade futbalu, kde sa výskum zameriava na 14 krajín, medzi ktorými sa nachádza napríklad aj Slovensko a Česká republika, v kategóriách od U8 (476296) po U12 (639153) dochádza k **34 % rastu** členskej základne mladých športovcov. U12 predstavuje najsilnejšiu kategóriu, kde **vyššie vekové kategórie dosahujú pokles**, U14 (565826) pokles **12 %**, U16 (470993) **pokles 17 %** a nakoniec U18 (523121), ktorá dosahuje **11 % rast** oproti kategórii U16. (EMMONDS, S. a kol. 2021)



Obrázok 3.1.1 Vývoj vekových kategórií vo výskume ICOACHKIDS – futbal
Zdroj: (EMMONDS, S. a kol. 2021)

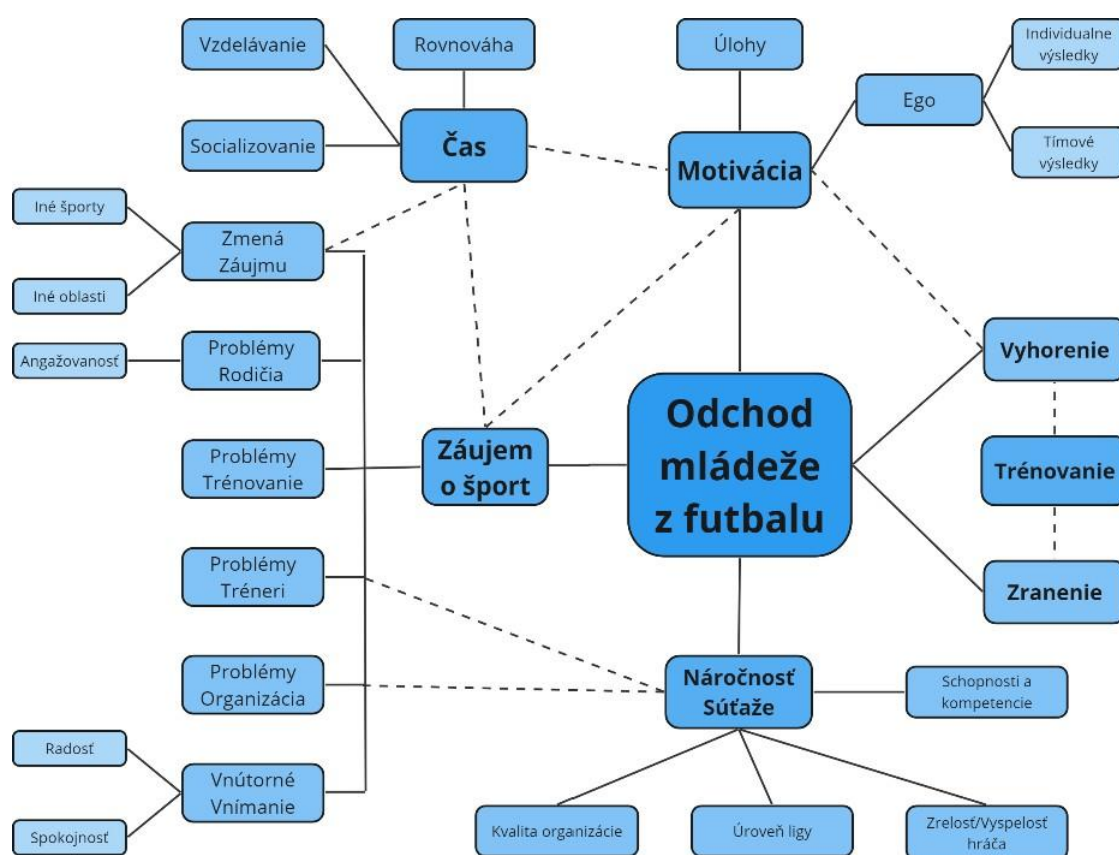
Dôležité je poznamenať, že pre výskum - odchod z organizovaného športu predstavuje ukončenie vykonávania hodnoteného športu. To znamená, že v prípade, ak **dieťa vykonáva dva alebo viac športov** a rozhodne sa **špecializovať iba na jeden, v ostatných športoch je chápané, ako keby skončilo so športovaním úplne**. Takýto druh odchodu predstavuje limitáciu výskumu. (EMMONDS, S. a kol. 2021)

Výsledky výskumu ukazujú, že všeobecná športová politika v Európe a špecifické športové politiky jednotlivých krajín sa musia zaoberať **významným poklesom mužov v mládežníckych kategóriách**. Vo väčšine športov k najväčšiemu poklesu dochádza u športovcov, ktorí prechádzajú z **vekových kategórií U14 a U18**. Toto obdobie je špecifické prechodom zo základnej školy na strednú školu a zo strednej školy na vysokú školu, prípadne do zamestnania. Pre zvýšenie výpovednú hodnotu je potrebné zistiť, ktorí športovci končia so športom úplne, a ktorí sa špecializujú na konkrétny šport. Aktuálne

neexistuje **jedinečný identifikátor športovcov** v jednotlivých športoch, a preto nie je možné sledovať migráciu detí a mládeže medzi jednotlivými športami. (EMMONDS, S. a kol. 2021)

3.2 Faktory vplývajúce na odchod z organizovaného športu

V nasledujúcej časti práce sa pojednáva o hlavných faktoroch, ktoré majú za následok odchod detí a mládeže z organizovaného športu. Za odchod zo športu u detí a mládeže do 15. roku života môže viacero príčin. Vo väčšine prípadov sa **faktory nevyskytujú samostatne, ale pracujú synergicky**. To znamená, že o skočení sa dá hovoriť ako o **multidimenzionálnom fenoméne**, kde medzi sebou vysvetľujúce faktory **korelujú**. (SCHLESINGER, T. a kol. 2018) Faktory popísané v nasledujúcej časti majú podľa výskumu NAYS (National Alliance of Youth Sports) za **následok skončenie viac ako 70 % detí** v organizovanom športe pred dovŕšením 13. roku života. Na **Obrázok 3.2.1** sú znázornené hlavné faktory, ich atribúty a niektoré vzťahy medzi faktormi navzájom.



Obrázok 3.2.1 Myšlienková mapa faktorov na základe analýzy literatúry
Zdroj: *vlastné spracovanie na základe podkapitol kapitoly 3.2*

Z Obrázok 3.2.1 možno vidieť, že na odchod z futbalu vplýva viacero faktorov, ktoré sú **často** alebo **úplne spolu prepojené**. Jednotlivé faktory možno definovať aj pozitívne, ale pre **účel práce sa na faktory pozerá z negatívnej strany**.

3.2.1 Záujem o šport

Dospievanie predstavuje obdobie, počas ktorého deti a mládež prechádzajú veľkým množstvom zmien. K najväčším zmenám v priebehu dospievania dochádza z hľadiska vývoja **fyzickej, emocionálnej a behaviorálnej stránky mládeže**. Počas transformácie detí a mládeže na dospelých jedincov dochádza k **formovaniu ich pocitu identity**, to znamená chápanie toho, kto sú a ako by ich mali ostatní vnímať. (BIGELOW,

B. 2016) V priebehu transformácie tak môže dôjsť k **zmene záujmov a preferencií** v oblastiach organizovaného športu ale aj mimo neho. V dotazníkoch zameraných na príčiny skončenia detí a mládeže s organizovaným športom sa ako najčastejšie dôvody vyskytujú odpovede typu: „*Mať iné veci na práci*“, „*iné sociálne priority*“ alebo „*pokračovať v inom športe*“. (SCHLESINGER, T. a kol. 2018)

Zaujímavé zistenia priniesol výskum inštitúcie **ISYS** (The Institute for the Study of Youth Sports), ktorý je vykonaný na vzorke viac ako 8000 mládežníckych športovcov, kde účelom je vytvorenie rebríčka **desiatich najdôležitejších dôvodov, prečo deti a mládež športujú**. U chlapcov sa **víťazstvo** ako dôvod pre športovanie umiestnilo až na **8. mieste** a u dievčat sa nedostalo ani do rebríčku (12. miesto). Výsledok výskumu ukázal, že ďaleko dôležitejšími faktormi, ktoré deti vnímajú, sú: **zábava** (prvé miesto aju chlapcov aj u dievčat), **zlepšenie schopností**, **udržanie si postavy** alebo **realizácia dieťaťa v oblasti, v ktorej je dobré**.

Zábava pri športe a radosť z jeho vykonávania predstavuje *jeden z kľúčových determinantov pokračovania* detí a mládeže v organizovaných športoch, vrátane futbalu. Výskum, ktorý používa logistickú regresiu ako nástroj na predikovanie pokračovania detí a mládeže s organizovaným futbalom, sa zameriava na skúmanie krátkeho (6 mesiacov) a dlhého obdobia (4 roky). Výsledky výskumu ukazujú, že okrem **radosti zo športu** vnímali „odídenci“ **viac alternatív**, čo sa týka iných športov, mali menej rodinných príslušníkov zainteresovaných vo futbale a **začali neskôr s futbalom ako športovci, ktorí pokračovali**. (VAN YPEREN, N. W. a kol. 2022)

Vplyv na záujem o šport u detí a mládeže je možné rozdeliť medzi nasledujúce **tri subjekty**, ktoré môžu najviac na záujem o organizovaný šport negatívne vplyvať:

Negatívny vplyv **rodičov** na záujem u detí o organizovaný šport sa prejavuje primárne, **nezvládaním emócií** vyplývajúcich z vývoja hry, ktoré vedú k **negatívnemu interagovaniu** (kričaniu, nadávaníu, agresivite a pod.) s hráčmi, rozhodcami alebo medzi rodičmi navzájom. Tento jav sa prejavuje v **negatívnom chápaní športového prostredia** u detí a mládeže, namiesto toho, aby mali športový výkon spojený s pozitívnou energiou. Ďalším javom, ktorý sa vyskytuje často a má zlý vplyv na záujem o organizovaný šport u detí, je rodičovské porovnávanie **detí medzi sebou a nadmerné chválenie vlastných detí** zo strany rodičov. Zmiený jav vytvára **tlak na deti**, kde mladí športovci sa viac zameriavajú na podanie kvalitných výkonov, čo môže vplyvať na zníženie vnímanej radosti zo športu v dôsledku transformácie voľnočasovej aktivity na povinnosť. Tento jav exponenciálne rastie, v prípade **nedosiahnutia stanovených očakávaní** mladého športovca alebo rodiča. Čím je podujatie prestížnejšie, tým sa efekt a pravdepodobnosť výskytu zmienených negatívnych javov zvyšuje. (BIGELOW, B. 2016)

Tréneri pre deti a mládež v organizovanom športe predstavujú „druhých rodičov“ a sú mnohými dospelými vnímaní **ako vzor**. Z toho dôvodu ich negatívne správanie, ktoré sa prejavuje viacerými spôsobmi, výrazne ovplyvňuje záujem o šport u detí a mládeže. Vlastnosť trénerov, ktorá je pri práci s dospelými chápaná ako najnegatívnejšia, je **túžba vyhrávať za každú cenu** aj v nižších vekových kategóriách. So spomenutou vlastnosťou úzko súvisí aj **posúvanie záujmov hráčov za záujmy trénera**, ktoré sa prejavuje vytváraním **nevyváženej zostavy** orientovanej na dosiahnutie najlepších výsledkov, čoho dôsledkom môže byť **nedostatok príležitostí** pre časť hráčskej základne alebo aj prelaženie niektorých športovcov. **Nadmerné hranie** predstavuje jav, kedy mladý športovec má vyťažovaný harmonogram tréningov a zápasov po celý rok, čoho výsledkom je **skrátene voľného času** na ostatné aktivity, medzi ktoré patria: iné športy, **vzdelávanie** alebo **socializovanie**. Spomenutý jav vedie k **preťaženiu/pretrénovaniu** mladého športovca, za ktoré môže primárne tréner. Výsledkom dlhodobého preťaženia je zvýšenie pravdepodobnosti **rizík** týkajúcich sa **fyzického a psychického zdravia športovca**. (BIGELOW, B. 2016) Nesprávne

navrhnutie harmonogramu športovcov a negatívne správanie sa trénera **zvyšujú pravdepodobnosť odchodu** detí a mládeže z organizovaného futbalu. (SCHLESINGER, T. a kol. 2018)

Športovci pôsobia v **organizáciách**, ktoré majú jeden alebo viac tímov umiestnených v rôznych vekových kategóriách. Problémy v elitných tímoch vznikajú primárne v dôsledku fyzických a emocionálnych požiadaviek na dospievajúceho. Často vyskytujúcou sa chybou zo strany organizácie je **skorá úzka špecializácia hráčov**, ktorá je rozobraná viac v časti *Vyhorenie*. Orientácia na výsledky u trénerov už v mladých vekových kategóriách je v mnohých prípadoch vyžadovaná organizáciou, kde ako **nástroj pre dosiahnutie cieľov organizácia využíva negatívne stimuly pre trénerov**. (BIGELOW, B. 2016) Počas obdobia, kedy je dieťa v mladom veku výber organizácie, v ktorej bude pôsobiť je hlavne na rodičoch. Za hlavné kritérium výberu organizácie pre svoje dieťa považujú rodičia **renomé organizácie**. Medzi najčastejšie problémy s renomé organizácie patria: **problémy s infraštruktúrou** (dostupnosť zariadení, kvalita vybavenia a pod.) alebo **problémy so zamestnancami** (nekvalifikovanosť, nekompetentnosť alebo nedostatok personálu). (SCHLESINGER, T. a kol. 2018)

V literatúre je organizácia chápaná aj ako združenie, zväz, ktorý zoskupuje viacero tímov z rôznych organizácií. V tomto zmysle je organizácia zodpovedná za vytvorenie majstrovských a vyradovacích súťaží, ktoré predstavuje faktor, na ktorý sa dá pozeriť z dvoch perspektív. Prvá predstavuje **vyššiu špecializáciu a skorú selekciu športovcov**, ktorí dosahujú nadpriemerné výkony vo svojej vekovej kategórii. Tým sa dokáže lepšie pracovať s talentom mladých športovcov, ale na druhej strane vytvára už v mladom veku u **detí tlak na podávanie kvalitných výkonov a dosahovanie úspechov**. (BIGELOW, B. 2016)

Medzi opatrenia zamerané na zachovanie záujmu u mládeže a zabráneniu odchodu z organizovaného futbalu v krátkodobom, ale aj dlhodobom horizonte patria: **zapojenie detí do futbalu vo veľmi mladom veku** (od 6 rokov) a **angažovanie súrodencov, rodičov a ďalších členov rodiny** v organizovanom futbale. (VAN YPEREN, N. W. a kol. 2022)

3.2.2 Náročnosť súťaže

Dospievanie športovcov **koreluje s rastúcimi požiadavkami** na fyzickú a psychickú vyspelosť. Proces transformácie športovcov do vyšších vekovými kategóriami vyžaduje zintenzívnenie tréningu a prípravy pre dosiahnutie úspechu. Niektorí športovci preto začnú vnímať šport viac ako **aktivitu podobnú práci** a nie ako predtým, hobby alebo koníček. To môže viesť u mladých športovcov k strate záujmu o šport.

V mladších ročníkoch sa počas transformácie môže okrem iného prejavíť faktor narodenia, kde **športovci narodení na začiatku roka** (v januári) majú výhodu oproti hráčom narodeným na konci roka. Získané **fyzické a psychické výhody oproti mladším** rovesníkom v kategórii im tak dávajú rôzne príležitosti, ktoré zvyšujú pravdepodobnosť úspechu v mládežníckych, ale aj dospelých kategóriách. V období dospievania je nárast v podobe 3 - 9 mesiacov výrazný. Vo futbale sa tento jav prejaví primárne v **kategórii žiakov**. (The Soccer Sidelines [online] 2019) **Rast/zrelosť** predstavuje predovšetkým nevýhodu z dôvodu zoskupovania hráčov do súťaží podľa veku. V futbalových kluboch, kde sú dva ročníky v jednej kategórii, môžu vzniknúť rozdiely medzi hráčmi až 23 mesiacov, pričom **biologické rozdiely pri dospievajúcich deťoch behom puberty môžu byť až 5 rokov**. (SCHLESINGER, T. a kol. 2018)

Športové, **futbalové schopnosti a kompetencie** športovca vplyvajú na radosť zo športu. Štúdie ukazujú, že športovci, ktorí hodnotia svoje schopnosti ako horšie oproti rovesníkom, majú **vyššiu tendenciu odísť zo športu skôr**. Existuje taktiež silná

korelácia medzi subjektívnym (hráčovým) a **objektívnym vnímaním schopností** kompetencií. Motivácia prostredníctvom **dosahovania úspechov** a následným dosiahnutím kompetencií v športe, zvyšuje vnímanie schopností a kompetencií jednotlivcom, ako aj jeho **nadšenie zo športu**. (SCHLESINGER, T. a kol. 2018)

Futbalisti, ktorí *dosahujú dobré výsledky*, sa **nezoberajú skončením s futbalom** tak často ako tí, ktorí *majú horšie výsledky*. Toto tvrdenie sa týka výsledkov na osobnej úrovni hráča, ale taktiež aj tímovej. (SCHLESINGER, T. a kol. 2018)

3.2.3 Trénovanie

Tréning a príprava predstavuje **väčšiu a dôležitejšiu časť** športu v mládežníckych kategóriách ako zápasy. V prípade, že deti a mládež nemajú pozitívnu skúsenosť s trénerom alebo štýlom tréningu, automaticky sa začne u nich **znižovať záujem o šport**. Tento faktor významne ovplyvňuje spokojnosť a vnímanú zábavu dieťaťa počas vykonávania športu. (BARQUERO-RUIZ, C. 2021) V tejto práci sa na tréning pozerá v kontexte dvoch oblastí, a to **pretrénovania a ideálneho začiatku s tréningom**.

Viaceré štúdie považujú **ranný vek** ako ideálny pre začiatok so športom. Spájajú skorý začiatok so zvýšením pravdepodobnosti sa presadiť alebo dosahovať nadpriemerný výkon v rôznych oblastiach. Príkladom môžu byť **huslisti**, kde skvelí do 18 veku priemerné odohrali 7410 hodín, dobrí už len 5301 hodín a tí, ktorí učia, okolo 3420 hodín. Tí, ktorí **začínajú neskôr, zvyčajne dosahujú vyššiu mieru zranení** z dôvodu pretrénovania pri snahe dobehnúť rovesníkov, ktorí začali skôr. Záverom štúdie je, že zámerné cvičenie pozitívne koreluje s expertným výkonom, ale vysvetľuje **len zlomok rozptylu schopností** (okolo 18 % v prípade športu). (REIDER, B. 2017) Z toho dôvodu je výhodnejšie trénovať v rámci možností dieťa, aby nestratilo záujem a zároveň nedošlo k pretrénovaniu, čo vedie k zvýšeniu šance zranenia.

Niektoré štúdie zdôrazňujú, že **odstránením nadmerného tréningu** sa zvyšuje pravdepodobnosť dlhodobej aktívnej účasti v športe ako aj šance dosiahnutia elitnej úrovne v dospelosti. **Oddych od povinnosti v športe**, niekoľkokrát do roka umožňuje mládeži viesť **komplexný život**, čo posilňuje motiváciu k športu po návrate. Dĺžka oddychu pritom nemusí byť dlhá (stačí 1-2 týždne) a môže byť využitá na venovanie sa ľubovoľným aktivitám dieťaťa napríklad školským povinnostiam alebo budovaniu vzťahom - **socializovanie**. (MOESCH, K. a kol. 2012) Nie je však dobre, ak je prestávka od športu príliš dlhá. Príkladom môže byť vynechanie tréningu po dobu 1 mesiaca z rôznych dôvodov. Pri snahe dobehnúť zameškaný tréning prostredníctvom vytvorenia nabitého tréningového harmonogramu s nedostatočným množstvom oddychu, sa zvyšuje pravdepodobnosť zranenia alebo vyhorenia. (REIDER, B. 2017)

V období mladého veku a dospievania je viac dôležitá pre dieťa **diverzifikácia** a prechod športovými aktivitami ako **špecializácia** na konkrétny šport. Následne keď dieťa príde do puberty prejde do fázy výberu športu a až v **neskoršom veku sa úzko špecializujú tréningy**, na konkrétny šport. (REIDER, B. 2017) Ako potvrdenie tvrdenia môže slúžiť štúdia 243 elitných športovcov v dospelosti, kde do 15. roku mali elitní športovci v dospelosti **výrazne menej odtrénovaných hodín v konkrétnom športe** ako reprezentanti v rovnakej vekovej kategórii. Do 18. roku sa počet hodín približne vyrovnával a po dovŕšení 21. roku počet odtrénovaných hodín výrazne prevyšoval počet hodín vzorky reprezentantov do 15 roku. (MOESCH, K. a kol. 2012) Ideálnym stavom je špecializácia, ktorá je výsledkom prirodzenej súčasti dospievania a nie rozhodnutím rodiča. Vykonávanie viacerých športov môže pomôcť k vytvoreniu viacrozmerneho pohľadu pre mladého športovca na seba samého – **vlastnej identity**. (VEALEY, R. S. a kol. 2016)

3.2.4 Vyhorenie

Ako je spomenuté v časti *Trénovanie*, v prípade, ak deti a mládež hrajú celý rok iba futbal, môžu začať cítiť pocit vyhorenia a stratiť záujem o šport. **Vyhorenie** v mládežníckych športoch nie je niečo, čo rodičia predpokladajú u svojich detí. **Ambicióznosť** predstavuje jednu z hlavných charakteristík pre dosiahnutie úspechu, avšak u detí ambície a silná vášeň, ktorá je **úzkou špecializovaná na jednu oblasť**, môžu viesť k vyhoreniu. Jedným z následkov vyhorenia je **chápanie športu ako povinnosti** a nie ako koníček alebo voľnočasová aktivita. (BIGELOW, B. 2016)

V časti *Náročnosť súťaže* je bližšie vysvetlený súťaživejší charakter športu, ktorý prispieva k **väčšiemu tlaku** na dieťa, čo zvyšuje šancu vyhoriť. Tento jav môže eskalovať až do štádia, kde dieťa si vytvára **chronický stres**. (GOMES, A. R. a kol. 2017) Medzi hlavné zdroje spomínaného tlaku patria buď **priamo rodičia**, ktorí vytvárajú tlak na dieťa alebo **nepriamo rodina**, ktorá vytvára takzvaný **nepriamy tlak**. Nepriamy tlak predstavuje tlak, ktorý nie je vytváraný zámerne rodinou, ale je **vnímaný dieťaťom** v dôsledku viacerých faktorov, napríklad športovej histórie rodiny alebo jej angažovanosťou v športe. (MARTINEZ-ALVARADO, J.R. a kol. 2021)

Tlak zohráva kľúčovú úlohu pri vyhorení. Často sa prejavuje vo forme **úzkosti z výkonu** (na tímovej, ale aj osobnej úrovni) alebo **negatívneho vnímania** – stresu, ktorý vychádza z vnímania seba samého ako málo kompetentného a málo efektívneho v športovej rovine. **Potreba uspokojenia požiadaviek** rodičov, rodiny alebo trénerov je častým zdrojom tlaku. **Sociálny tlak** predstavuje tlak, ktorý pramení z vnímania rovesníkov. V **príbehu dospievania** sa tlak zo spomenutých subjektov mení, kde s rastúcim vekom športovca vplyv **tlaku rodiča klesá** a **tlak zo strany rovesníkov rastie**. (SCHLESINGER, T. a kol. 2018)

Medzi osobnostné črty mládeže, ktorá patrí do skupiny náchylnejších na vyhorenie, patria: **negatívny perfekcionizmus** (sklon k negatívnemu sebahodnoteniu a nespokojnosť, kde dosahovanie výsledkov je poháňané negatívnou motiváciou alebo strachom zo zlyhania), **úzkostné rysy** (nadmerné pociťovanie nepríjemných pocitov ako sú obavy a strach), **obsedantná vášeň** (nedokázanie myslieť na nič iné ako šport, abstrahovanie iných aktivít) s čím je spojená aj **jednorozmerná identita**. (BIGELOW, B. 2016), (GOMES, A. R. a kol. 2017)

Medzi časté príznaky **začiatku** ale **priebehu vyhorenia** u detí a mládeže patria:

- **Chronická únava** – v dôsledku vytvárania fyzických a psychických nárokov na deti a mládež subjektami. Bližšie nároky na dieťa a mládež charakterizujeme v časti *Záujem o šport*.
- **Devalvácia a odlúčenie** – strata záujmu o šport a jeho vykonávanie.
- **Zníženie pocitu úspechu** – negatívny pohľad na dosiahnuté výkony a úspechy (príkladom môže byť negatívny perfekcionizmus).
- **Faktory pôsobia navzájom a podmieňujú sa**, napríklad zvýšené **trénovanie** vedie k **vyčerpaniu** - únave, ktoré vedie k **devalvácii** a **postupnej strate záujmu** o šport. (BIGELOW, B. 2016), (GOMES, A. R. a kol. 2017), (MARTINEZ-ALVARADO, J.R. a kol. 2021)

Atribúty, ktoré zvyšujú riziko vyhorenia, sú: **hranie jedného športu po dobu viac ako 8 mesiacov v roku**, **nadmerné trénovania**, **neprimerané očakávania** na úspech, **nízke sebavedomie**, **tlak** so strany rodičov, trénerov alebo rovesníkov. (GOMES, A. R. a kol. 2017)

Pri vytváraní strategických programov za účelom **prevencie voči vyhoreniu** je dôležité zahrnúť ako aspekt **pozitívnu a harmonickú vášeň mladých športovcov**. Je potrebné držať sa v intervale, ktorý nepatrí do extrému, pretože negatívny

perfekcionizmus a obsedantná vášeň vplývajú na zvýšenú mieru výskytu vyhorenia. Toto tvrdenie podporuje aj štúdia 452 mexických športovcov vo veku od 12 do 18 rokov, ktorej výsledky ukazujú na **obsedantnú vášeň ako faktor zvýšenia rizika športového vyhorenia**, rozvíjania negatívnych pocitov voči športu, ale aj prežívaniu individuálnych negatívnych konfliktov mladého športovca. (MARTINEZ-ALVARADO, J.R. a kol. 2021)

Výsledkom dlhotrvajúceho vyhorenia je často-krát vytvorenie zmýšľania u detí a mládeže, kde úspech v športe predstavuje **zvýšenie noriem a požiadaviek**. To znamená, že **úspech sa stáva prakticky nedosiahnuteľným**, pretože zlepšením vlastných schopností a kompetencií sa vnímaná úroveň, ktorú musí mladý športovec dosiahnuť, zvýši. Tento jav vedie k strate záujmu u detí a mládeže o šport.

3.2.5 Čas

S dospievaním dieťaťa prichádza **rast povinností** a príchod iných možností. Jedná sa napríklad o brigády a iné zárobkové činnosti, zvyšujúca sa náročnosť školských úloh a iné pridružené povinnosti, ale aj väčšia zodpovednosť za domáce práce. Tieto atribúty vplývajú na voľný čas dieťaťa, ktoré sa môže rozhodnúť, že **voľný čas už nebude venovať športu, ale radšej iným aktivitám**.

Obmedzený čas môže predstavovať dôvod pre skončenie so športom pri mladých športovcoch. **Znížené množstvo času vplýva na početnosť a dĺžku tréningov**, respektíve prípravu na zápasy alebo súťaže. Nedostačená príprava môže viesť k **poklesu výkonnosti** – dosahovaniu horších športových výsledkov alebo k snahe o dobehnutie zameškanej prípravy zvyšovaním intenzity a nárazovej záťaže počas tréningu – **pretrénovanie**, čo zvyšuje riziko **zranenia**. Samotné zníženie času vykonávania športovej aktivity môže znížiť záujem o šport u dieťaťa. (MERKEL, L. D. 2013)

Samostatný atribút vplývajúci na čas dieťaťa počas dospievania je **vzdelávanie**. Prechádzaním do vyšších stupňov vzdelávania počas dospievania sa priamo úmerne zvyšuje náročnosť dosahovania dobrých výsledkov. Tento faktor môže viesť k **priorizovaniu štúdia nad šport** zo strany dieťaťa alebo rodiny. **Akademický tlak** tiež zvyšuje pravdepodobnosť vzniku problémov s duševným zdravím, ako je napríklad **stres** alebo **úzkosť**. To môže viesť k vyhoreniu, zníženiu radosti zo športu, zníženiu pocitu úspechu alebo straty záujmu o šport. Dospievajúci športovci môžu vnímať počas postupu do vyšších ročníkov zvýšený tlak školy a tréningov, čo vedie k únave a pretrénovaniu. (DOS SANTOS, L. M. a kol. 2020)

Štruktúrálnou podmienkou určujúcou pokračovanie športovca v organizovanom futbale patrí aj **nájdenie rovnováhy** medzi časom stráveným pri **vykonávaní futbalu** (potrebný čas na tréningy, prípravu, zápasy a pod.) a **voľnočasovými aktivitami** (socializovanie, iné športy, vzdelávanie alebo pod.). (SCHLESINGER, T. a kol. 2018)

3.2.6 Ostatné faktory

Ostatné faktory predstavujú skupinu faktorov, ktoré majú minimálny alebo nízky vplyv na skončenie s organizovaným futbalom na Slovensku. **Nedostatok príležitostí**, ktorý vzniká v dôsledku problému s infraštruktúrou, kde na Slovensku vo futbale nie je tento problém a z toho dôvodu je zaradená v tejto časti. Rovnako **náklady** zohrávajúv niektorých športoch ako napríklad hokej dôležitý faktor, ale pre futbal nepredstavujú zásadný problém.

Nedostatok príležitostí predstavuje v niektorých oblastiach športu problém, kde deti majú záujem o vykonávanie športu, ale nie je vytvorený dostatočný počet príležitostí pre vykonávanie športu. Hlavným atribútom odstránenia faktoru nedostatočnej príležitosti je vybudovanie dostatočnej **infraštruktúry pre šport**. Konkrétne podľa

zákona o športe sa dá športová infraštruktúra definovať ako športová hala, športové ihrisko, telocvičňa a iné kryté alebo otvorené športové zariadenie určené na vykonávanie športu (§ 3, pís. o, Zákona č. 440/2015 Z.z.). Na Slovensku stav športovej infraštruktúry nie je ideálny, kde počet športovísk na 1000 obyvateľov je **0,84**, kde napríklad Dánsko (vybrané na základe približne rovnakého počtu obyvateľov 5,7 mil. ako Slovensko – 5,5 mil.) má hodnotu metriky na úrovni **2,07**. (STATISTICS DENMARK [online] 2022) Najviac rozšírenou športovou infraštruktúrou na Slovensku je práve futbal a preto nepredstavuje nedostatok príležitostí zásadný faktor. (MŠVVaŠ SR, 2022)

Futbal nepatrí medzi športy, ktoré sú nákladné na vykonávanie, ale stále môže byť pre niektoré rodiny menej dostupný. Financie predstavujú veľmi slabý faktor, pretože vo väčšine líg sú **poplatky klubu minimálne**. Jedinou oblasťou, ktorú financie ovplyvňujú, je **dostupnosť futbalových kempov** ako nástroja na zlepšenie a dosahovanie výnimočných výsledkov. (SCHLESINGER, T. a kol. 2018) S **nákladmi** sú spojené aj vedľajšie parametre napríklad **viac-dimenzionálnosť mladého športovca** ako osoby, kde deti bohatších rodín vyskúšali v priemere viac športov ako deti z nižšieja strednej triedy. Bohatšie deti taktiež začínajú a končia so športom skôr, vnímajú menej tlaku na sebe, majú viac zameraných tréningov a ich rodičia viac minú na výdavky okolo športu (lepšie vybavenie, tréningy a pod.). Výsledkom spomenutých parametrov je, že deti z rodín s nižším príjmom majú 6-násobne väčšiu pravdepodobnosť skončenia so športom v dôsledku nákladov. (SOLOMON, J. 2020)

Rodičia, hlavne v rodinách s nie veľkým príjmom, častejšie dávajú **viac povinností a zodpovedností na deti** (starostlivosť o súrodencov alebo starších členov rodiny). Deti a mládež, ktoré sú týmto spôsobom angažované v rodinných úlohách, môžu obmedziť šport, prípadne úplne **stratiť záujem o šport v dôsledku rodinných priorít**. (BIGELOW, B. 2016)

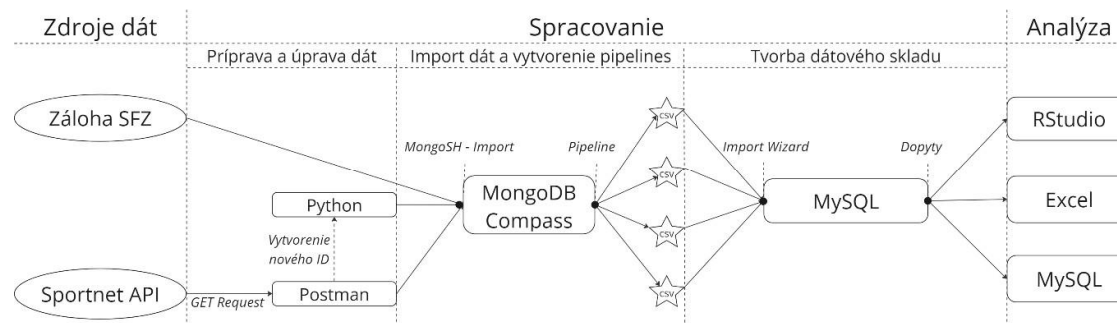
3.2.7 Zhrnutie faktorov vplyvujúcich na odchod

Analýza faktorov, vplyvujúcich na odchod mládeže z organizovaného športu po dovŕšení 15. roku, odhalila viacero informácií. Faktory, ktoré vysvetľujú viac ako 70 % odchodu mládeže zo športu, sú primárne zamerané na **vnútorné prežívanie športovca**. Skoro všetky faktory pracujú s **prepojením vnímaných pocitov športovca** ako sú *radosť, spokojnosť a zábava* so športom. Subjekty, ktoré *vplývajú na dieťa najviac (rodičia a tréneri)*, sú zodpovedné za vytvorenie a udržiavanie týchto pocitov. V mnohých prípadoch však **vytvárajú negatívne emócie a pocity u mládeže**, medzi ktoré patria *tlak, stres, úzkosť* alebo *strach*. *Kvalitu, frekvenciu a zábavu tréningovania* vnímanú mládežou po fyzickej, ale aj psychickej stránke, literatúra považuje za kľúčový faktor. Nedostatky v tréningu sa prejavujú negatívne na mládeži vo forme *zranenia, pretrénovania, vyhorenia, strate záujmu, jedno-dimenzionálnej identity športovca*. Fyzickú a psychickú výhodu môže športovec získať oproti rovesníkom aj v podobe **mesiacu narodenia**. Jedná sa o športovcov, ktorí sa narodili na začiatku roka (primárne prvý kvartál). Viaceré štúdie sa zhodujú, že v mládežníckom športe by viac ako úspech mal byť dôležitý **rozvoj, rovnováha a radosť** mládeže. Preto je dôležité poznamenať, že odchod z organizovaného športu predstavuje jav, ktorý je **prirodzený** a účelom by nemalo byť dosiahnutie *celkového odstránenia odchodu*, ale **optimalizovanie negatívnych podmienok vplyvujúcich na odchod** ako nekvalitné tréningovanie, tlak a ďalšie.

Pre ďalšie pokračovanie v práci analýza faktorov ukázala, že dáta, ktoré sú potrebné na predikovanie odchodu z organizovaného športu, sú v štúdiách **získavané dotazníkovým prieskumom**. Dôvodom výberu tejto metódy je zameranie viacerých faktorov na **interpersonálnu oblasť športovca**, kde dáta, ktoré sú zaznamenané SFZ, sú nepostačujúce na predikovanie odchodu. Dáta, ktorú sú dostupné, vyjadrujú primárne len jeden faktor, a to **úspech** (tímový aj individuálny) a sekundárne **motiváciu zameranú na ego**.

3.3 Spracovanie dát

V nasledujúcej časti sa rozoberá spôsob spracovania údajov a transformácia dát do vhodnej podoby pre *získavanie informácií, vedomostí a znalostí z dát*. Dáta sa tiež agregujú vo viacerých dimenziách na poskytnutie prehľadu – informačného panelu. Dáta, s ktorými sa primárne zaoberá v práci, sa dajú rozdeliť na dva hlavné zdroje, a to dáta získané z **verejne dostupných API sportnet** a **zo zálohy databázy Slovenského futbalového zväzu**. Obrázok 3.3.1 predstavuje ilustráciu procesu spracovania dát, kde sú zobrazené fázy, zdroje dát a využité softvérové nástroje.



Obrázok 3.3.1 Diagram zdrojov, nástrojov a fáz spracovania dát
Zdroj: *vlastné spracovanie*

Zdroj dát s názvom **Záloha SFZ** obsahuje v objektovej podobe (JSON - array) zápisy zo stretnutí, ktoré sú zastrešované Slovenským futbalovým zväzom od sezóny 2012/2013. Objekt je následne zložený s dátových typov, akými sú number, string, boolean alebo aj object a array, ktoré v sebe obsahujú ďalšie **vnorené objekty** alebo **listy s údajmi**. Ďalším zdrojom dát sú **verejne dostupné API** (API sportnet, 2023), kde sa nachádzajú ďalšie doplnujúce údaje o hráčoch, kluboch alebo organizáciách. V práci sa používajú API sportnet s názvami **Transfer** a **User**.

3.3.1 Príprava dát na analýzu

Na získanie dát z API sa posielali volanie na server **sportnet** (GET Request) v prostredí **Postman** (platforma API slúžiaca pre navrhovanie, zostavovanie, testovanie a opakovanie API). Týmto spôsobom sú získané dáta **Transfer** a **User** (detailnejšie popísanie týchto tabuliek/dokumentov neskôr). Pred importom dát z API Transfer, dáta museli prejsť procesom **vytvorenia unikátneho identifikátora ID** (zloženého kľúča), pretože pri pokuse o import dochádzalo k chybe z dôvodu duplicitných ID dokumentov. Tento problém vznikol ako výsledok, snahy vloženia rovnakého ID viac-krát. V zápise o transferoch slúži identifikátor ID hráča, kde jeden hráč môže byť súčasťou transferu (prestup/hosľovanie) **viackrát počas kariéry**. Pre vytvorenie **zloženého kľúča** je použitý softvérový nástroj **Python**, ktorý cyklicky prechádza každý dokument, kde mení `_id` (jedinečného identifikátora) na `player_id` a pridáva nový string, ktorý predstavuje `player_id + date_from` (dátum začiatku transferu), ktorý priradí ako nové `_id` zápisu o transfery. Týmto spôsobom sa vyriešil problém duplicitných vstupov, pretože hráč nemôže prestúpiť v jeden čas viac-krát.

Pre získanie potrebných údajov a informácií z dostupných zdrojov, nakoľko sa jedná o súbor typu JSON, je nutné použiť nástroj na prácu so súborom JSON alebo jeho **parcing** (proces premeny JSON objektov na textový formát). Rozhodlo sa pre možnosť, kde **dáta pred analýzou budú konvertované z neštruktúrovanej podoby do štruktúrovanej**. Ako prvý pokus tohto procesu sa využíval programovací nástroj Python, ktorý však nedokázal načítať dáta do operačnej pamäte (súbor zálohy má viac ako 10 GB) a následným spracovaním do požadovanej podoby cez **knižnicu JSON**. Z toho dôvodu v ďalšom pokuse sa použila **knižnica IJSON**, ktorá znižuje pamäťovú náročnosť spracovaním súboru cez **iterácie** (opakovanie toho istého procesu s parciálnymi

časťami súboru). Tento proces taktiež zlyhal. Na základe predchádzajúcich pokusov sa spracovanie dát nakoniec vykonáva primárne na úrovni dokumentovej databázy typu **NoSQL**, z ktorého pochádza aj záloha, a to **MongoDB**. Konkrétne sa používajú softvérové nástroje **MongoDB Compass** a **mongosh**.

MongoDB Compass predstavuje **silné GUI** (grafické rozhranie operačného systému, ktoré využíva ikony, rôzne ponuky, nástroje alebo kurzor na riadenie interakcie so systémom) na **dopytovanie**, **sumarizovanie** (agregovanie) a **analyzovanie dát** z NoSQL databázy MongoDB vo vizuálnom prostredí. (MongoDB Docs Compass [online] 2023) Prostredie umožňuje vytvorenie **dátových kanálov** (data pipelines), ktoré dokážu **rozbaľiť** a **izolovať objekty** do potrebnej miery a následne vytiahnuť požadované informácie prostredníctvom **vytvorenia jednotlivých etáp dátového kanálu**. Pre transformáciu dát do požadovanej formy sú v prostredí MongoDB Compass využívané najmä tieto etapy dátového kanálu (pipeline):

- **group** – zoskupenie dokumentov podľa definovaného id – špecifického výrazu zloženého z časti objektu („stpcov“),
- **match** – filtrovanie dokumentov podľa špecifickej podmienky alebo pod-dotazu („subquery“),
- **project** – vytvorenie novej časti dokumentu s vypočítanou hodnotou, zmenenie poradia častí dokumentu alebo špecifikovanie polí, ktoré sa majú zobrazíť vo výstupe,
- **sort** – zoradenie dokumentov podľa špecifického kľúču a zvoleného smeru (ASC, DESC),
- **unwind** – rozdelenie elementov v liste (array) v jednom dokumente na samostatné záznamy vo výstupe. (MongoDB Docs Compass 2023)

Dátové kanály (pipelines) majú podobnú štruktúru, kde sa začína v prípade potreby podmienkou na filtrovanie **match**. Následne sa rozbalia potrebné listy (array) v každom dokumente pomocou etapy **unwind**. Tento krok je dôležitý hlavne v prípade, ak sa vykonávajú agregované operácie na základe podmienky. Príkladom môže byť objekt „*protocol*“, ktorý obsahuje list „*events*“, kde sú zaznamenané jednotlivé udalosti zo zápasu (gól, asistencia a ďalšie). Pri počítaní výskytu je potrebné, aby každá udalosť predstavovala samostatný dokument. Takto pripravené dáta sa následne môžu agregovať (sumarizovanie dát na základe zvolených atribútov) prostredníctvom etapy **group**, kde v prípade potreby sa vykonávajú aj kalkulácie. V závere sa na zobrazenie potrebných údajov v dátovom kanáli využije etapa **project**. Príklad, ako vyzerá skript dátového kanálu, je na *Obrázok 3.3.2*.

```
[
  {
    $unwind: {
      path: "$protocol",
    },
  },
  {
    $unwind: {
      path: "$protocol.events",
    },
  },
  {
    $group: {
      _id: {
        season: "$season._id",
        competition: "competition._id",
        competitionPart: "$competitionPart._id",
        player_id: "$protocol.events.player._id",
        event_type: "$protocol.events.eventType",
      },
      count_event_sccP: {
        $count: {},
      },
    },
  },
]
```

Obrázok 3.3.2 Ukážka skriptu dátového kanálu v MongoDB CompassZdroj: *vlastné spracovanie*

Na Obrázok 3.3.3 je zobrazený jeden výstup, ktorý reprezentuje jeden riadok v relačnej podobe.

```
{
  "_id": {
    "season": {
      "$oid": "5e2ed5a1dabb0396ccc6c08c"
    },
    "competition": "competition._id",
    "competitionPart": {
      "$oid": "5e2fff337327b94aeb867c80"
    },
    "player_id": "5d65878c86dc8b723832be3f",
    "event_type": "goal"
  },
  "count_event_sccP": 25
}
```

Obrázok 3.3.3 Ukážka výstupu dátového kanálu v MongoDB CompassZdroj: *vlastné spracovanie*

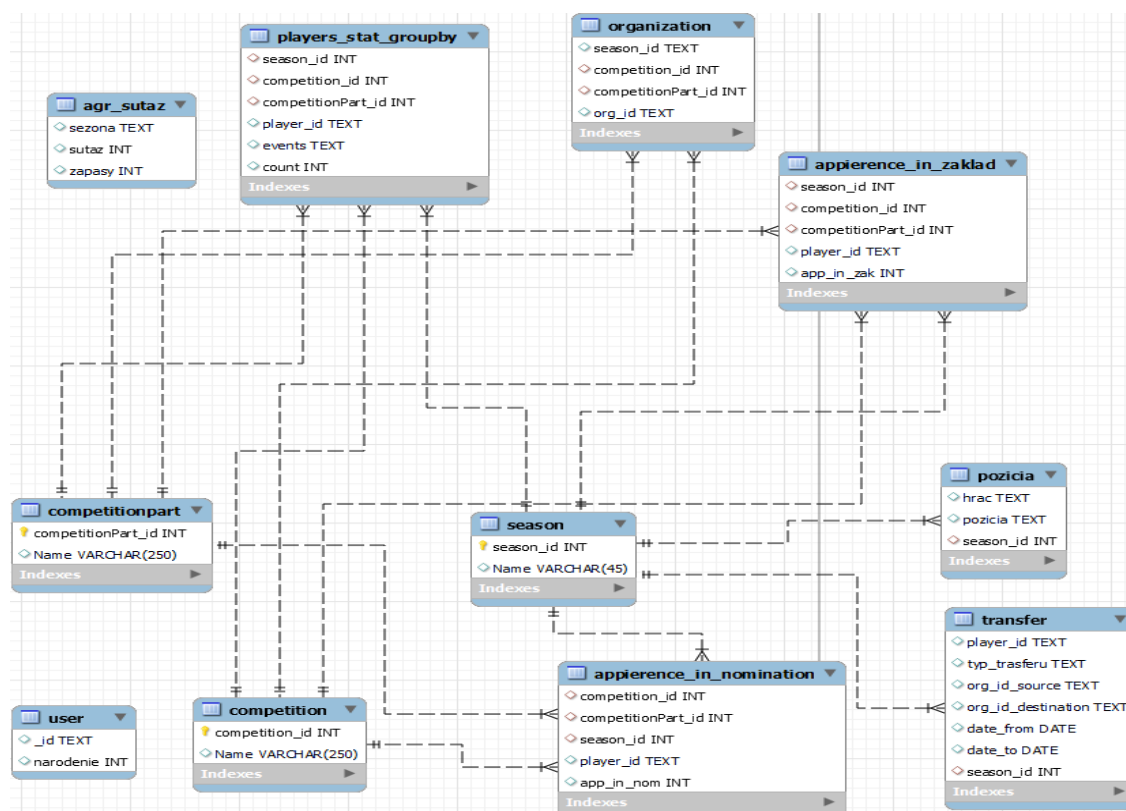
Jednotlivé **dátové kanály** sa následne v procese **exportu** transformovali na záznamy v CSV súbore a nahrali v **importe** do dátového skladu v MySQL databáze.

3.3.2 Vytvorenie dátového skladu

Po vytvorení dátových kanálov (data pipelines) sa dáta vyexportovali do súborov typu **CSV**. CSV súbory sa následne nahrávajú do **MySQL**, kde každá tabuľka predstavuje samostatný dátový kanál. Niektoré operácie sa dajú vykonať na úrovni **MongoDB Compass súbežne**, ale z dôvodu zachovania **integrity spracovania dát** jednotlivých krokov v jednom prostredí a väčších skúseností s vytváraním dopytov v MySQL sa použilo prostredie MySQL na dokončenie. Pred zobrazením vytvoreného **ERD** (subjektovo-vzťahového) **modelu** sa definujú základné vlastnosti spracovaného súboru dát.

Jedinečný identifikátor hráča sa nachádza v stĺpci `player_id` (napríklad 5d65766b86dc8b72382f0c35). Spracované dáta je potrebné vnímať **v časovom horizonte**, kde číselník pre časový horizont – sezónu sa nachádza v tabuľke `season`. Pre účel práce je nevyhnuté dáta zoskupovať podľa `season_id`, ktoré predstavuje číselný identifikátor sezóny (napr. id - 11; name - 2022/2023). Ako ďalšie sumarizačné kľúče a **nástroje filtrovania** vybranej kategórie športovcov sú použité `competition_id` (špecifikovanie kategórie a geografického umiestnenia napr. I. LMŽ - U13 liga mladší žiaci do 13 rokov) a `competitionPart_id` (časť súťaže napr. Základná časť).

Na Obrázok 3.3.4 je zobrazený **ERD model**, kde je vidieť okrem vzťahov aj jednotlivé stĺpce, respektíve časti v dokumente.



Obrázok 3.3.4 Vzor ERD diagramu dátového skladu

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

- **appierence_in_nomination** – zdroj údajov predstavuje záloha poskytnutá SFZ, ktorá je od sezóny 2012/2013. Všetky stĺpce okrem `app_in_nom` slúžia ako zložený kľúč pri vytváraní **dátového jazera**.
- **appierence_in_zaklad** – rovnaký zdroj a štruktúra ako `appierence_in_nomination` s rozdielom, že sumárny údaj predstavuje stĺpec `app_in_zak`.
- **player_stat_groupby** – predstavuje agregovanú tabuľku na základe zloženého kľúča, ktorý vyjadruje **typ udalosti** (striedanie, gól, asistencia, žltá karta, druhá žltá karta, červená karta a neuznaný gól) a k nemu príslušný počet výskytov udalosti podľa agregácie (sezóny, súťaže, časti súťaže a hráča).
- **competition** – získaná rovnako zo zálohy SFZ, poskytuje názov jednotlivých súťaží spolu s `competition_id`, ktorý predstavuje pomyselný primárny kľúč. V zálohe má názov `jssf`.
- **competitionPart** – rovnaký princíp získania dát a štruktúra ako `competition`.

- **pozícia** – zdroj dát je získaný rovnako ako predchádzajúce tabuľky zo zálohy, kde hlavnú informačnú hodnotu nesie stĺpec **pozícia**. Stĺpce **season_id** a **hrac** predstavujú zložený kľúč. Hráč môže byť na viacerých pozíciách počas jednej sezóny. Možné pozície sú: goalkeeper, defender, midfielder alebo forward.
- **season** – tabuľka sezóna je vytvorená z dôvodu zrýchlenia vyhľadávania sezóna a nahradenia textového prehľadávania, číselným. Tabuľka **season** môže byť označená aj ako číselník.
- **user** – zdroj dát pre tabuľku **user** predstavuje API od spoločnosti sportnet. Poskytuje viaceré informácie ale pre účel práce, ako relevantný je **vyhodnotený rok narodenia** a **dátum registrácie hráča**. Medzi ďalšie informácie, ktoré poskytuje API, patria napríklad ocenenie, ktoré sa poskytuje iba raz ročne, a preto má ako kritérium minimálnu výpočtovú hodnotu.
- **transfer** – kde zdrojom dát je taktiež API od spoločnosti sportnet. **Typ transferu** môže byť buď **hostovanie** alebo **prestup**. Taktiež API Transfer poskytovala viacero informácií, ale spracovania transferov a vyjadrenie migrácie medzi jednotlivými súťažami predstavuje problematiku samú o sebe, a preto pre účel tejto práce sa použili len vybrané údaje: dátum od, do, typ transakcie, z akej organizácie, do ktorej organizácie a následne zložený kľúč sezóny (**season_id**) a hráča (**player_id**)

Na vytvorenie finálneho súboru dát pre **trénováciu** a **testovaciu množinu** sa použilo prostredie MySQL, kde sa parciálne časti dát spájali prostredníctvom dopytu. Nakoľko sa však jednalo o spájanie 10 tabuliek (JOIN), kde niektoré tabuľky mali niekoľko desaťtisíc záznamov a **spájanie** (ON) prebiehalo na **základe stringu** (textového reťazca), odpoveď na dopyt niekedy neprišla. Prvým riešením problému je **CTE** (Common Table Expressions), ktoré umožňuje vytvorenie viackrokovej dátovej transformácie v lineárnej postupnosti. CTE riešenie aj keď poskytuje správny výsledok, nebolo optimálne z dôvodu náročnej údržby v prípade zmeny. Z toho dôvodu finálne riešenie pre zobrazenie údajov predstavuje **dočasná tabuľka**. Dočasná tabuľka umožňuje vytváranie indexov, taktiež znižuje náročnosť výpočtu z dôvodu obmedzenia dátovej základne, dokáže pracovať s premennými a umožňuje dopĺňanie údajov postupne, čo znamená aktualizovanie (UPDATE). V prípade zjednodušenia sa dá taktiež využiť ako black-box v procedúre, kde sa vložia len potrebné vstupné parametre a užívateľ dostane potrebný výsledok bez toho, aby musel viesť logiku dopytu.

Tabuľka 3.3.1 Finálne čistenie dát od odchýlok v MySQL

Sezóna	Počet		Vyradenie			Podiel vyradenia
	Celkový	Konečný	Ročník	Logika	Brankár	
2	3245	2740	9	140	372	4,62
3	3418	2887	9	145	391	4,62
4	3326	2855	3	126	354	5,21
5	3405	2840	6	131	384	5,49
6	3214	2770	5	107	341	3,58
7	3604	3099	1	119	397	3,36
8	3350	2886	1	88	386	2,63
9	3324	2800	12	69	350	2,57

Zdroj: *vlastné spracovanie*

V Tabuľka 3.3.1 sa nachádza vyradenia športovcov podľa sezón (2 = 2013/2014 až 8 = 2020/2021). Podiel vyradených hodnôt predstavujú maximálne 6 %, a preto dáta stále môžu byť považované za **neskreslené**.

Vzor výstupu dát po vykonaní transformácie je na **Obrázok 3.3.5**.

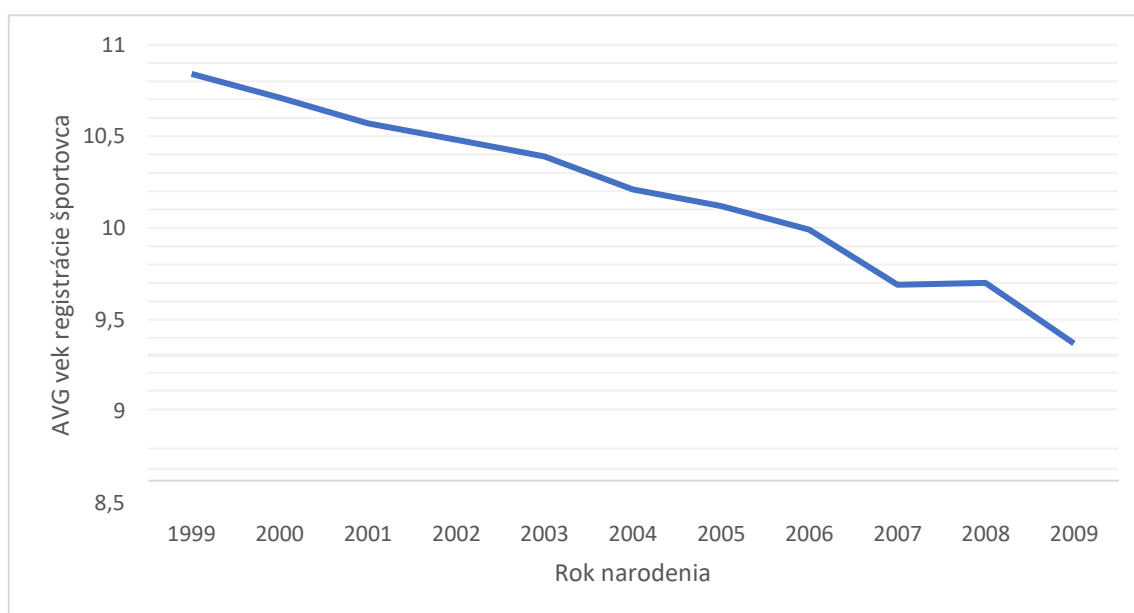
season_id	player_id	narodenie	old_when_start	app_in_nom	app_in_zak	Substitution	Goal	YellowCard	RedCard	Obranca	Zaloznik	Utocnik	viac_pozici	transfer_minulu	Pokracuje	norm_narodenie
9	5d658a6386dc8b7238334393	2006	8	15	4	1	0	0	0	0	0	1	0	0	0	0.0000
9	5d658c1e86dc8b7238338deb	2007	8	11	5	0	0	0	0	0	1	0	0	0	0	-0.2500
9	5d658dbb86dc8b723833d825	2006	10	11	7	1	1	1	0	1	0	0	0	0	1	0.0000
9	5d658afb86dc8b7238335ca6	2006	9	11	5	0	0	0	0	0	0	1	0	0	1	0.0000
9	5d65844d86dc8b723833229ed	2006	7	15	5	4	2	0	0	0	1	0	0	0	1	0.0000

Obrázok 3.3.5 Vzor výstupu agregácie SFZ zálohy v MySQL
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Pred vloženíím tréningových a testovacích dát do prostredia **RStudio** sa odfiltrovali hráči, ktorí sú na pozícii **Brankár** (goalkeeper), tí, ktorých **registrácia v SFZ** (old_when_start) je nižšia ako 5 a väčšia ako 14 (takýto hráči môžu skresľovať dáta) a tí, ktorých **rok narodenia je najväčší v sezóne**. Brankári sú vyradení z dôvodu štatistik, ktoré sa u nich sledujú.

3.4 Dolovania dát poskytnutých SFZ

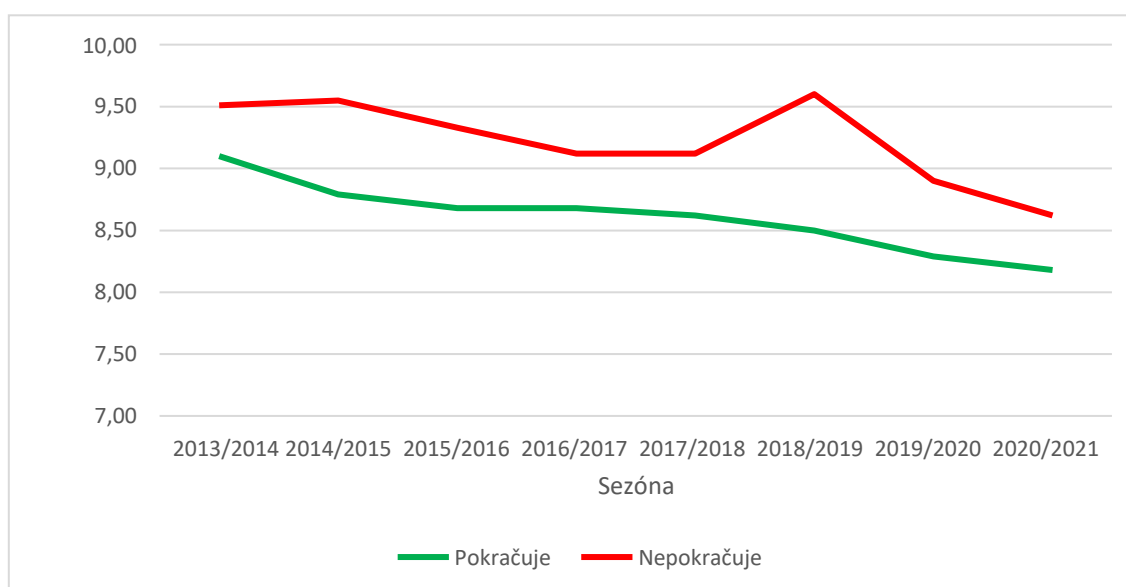
V podkapitole ťaženie dát sa primárne využívajú techniky **prieskumnej analýzy dát**, kde sa agregujú dáta a hľadajú sa rozdiely medzi dvoma kategóriami (pokračujúca nepokračujúcimi) a **techniky objavovania vzorov**, ktoré pomáhajú definovať trendy a vzory v správaní jednotlivých kategórií. Interpretácia zistení analýzy je formou zobrazenia grafu, popisu, kde na konci sa nachádza zhrnutie zistení z podkapitoly ťaženie dát.



Graf 3.4.1 Priemerný vek registrácie športovcov do SFZ - rok narodenia 1999 - 2009

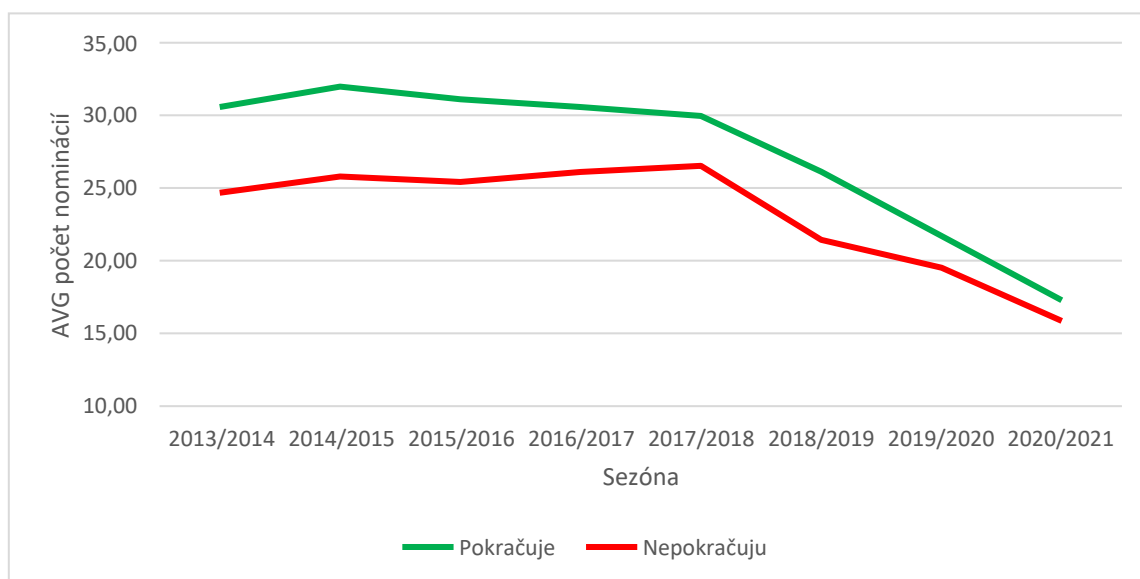
Zdroj: *vlastné spracovanie na základe API sportnet - User*

Graf 3.4.1 hovorí o **priemernom veku registrácie** športovca v slovenskom futbalovom zväze. Dáta sú získané z sportnet API, kde filter údajov predstavuje rok narodenia od 1999 po 2009 a rok registrácie, kde športovci, ktorí sú registrovaní pred dovŕšením 5. roku narodenia a tí, ktorí sú registrovaní po 14. roku narodenia sú vylúčení. **Štandardná odchýlka** registrácie športovcov je približne pre všetky roky na úrovni 2,03. Graf 3.4.1 ukazuje pozitívny trend, a to **poklesu veku, v ktorom sa športovci registrujú**, respektíve o registrácii môžeme hovoriť vo väčšine prípadov ako o **začiatku s organizovaným futbalom**.



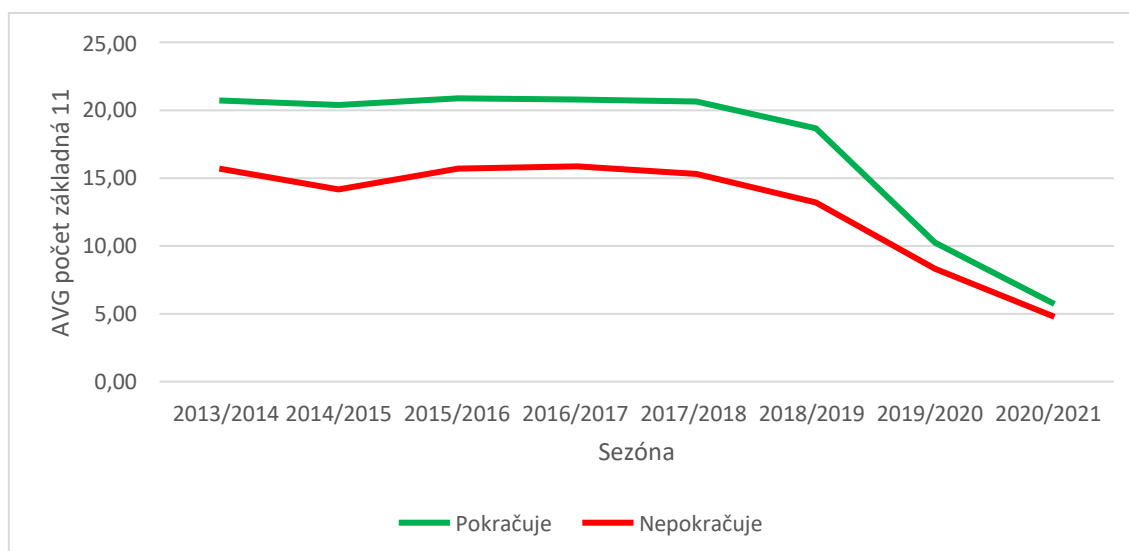
Graf 3.4.2 Priemerný vek žiakov pri registrácii za sezónu podľa závislej premennej
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet - User*

Graf 3.4.2 ukazuje na rozdiel medzi priemernou registráciou mladých športovcov do SFZ, ktorí pokračujú a tých, ktorí nepokračujú s organizovaným futbalom v zväze. Z Graf 3.4.2 je možné vidieť, že **vek hráčov, kedy sa registrujú, klesá** a druhým zistením je, že tí športovci, ktorí sa registrujú skôr, čo môže reprezentovať aj skorší začiatok s organizovaným futbalom, majú **väčšiu pravdepodobnosť, že budú pokračovať**.



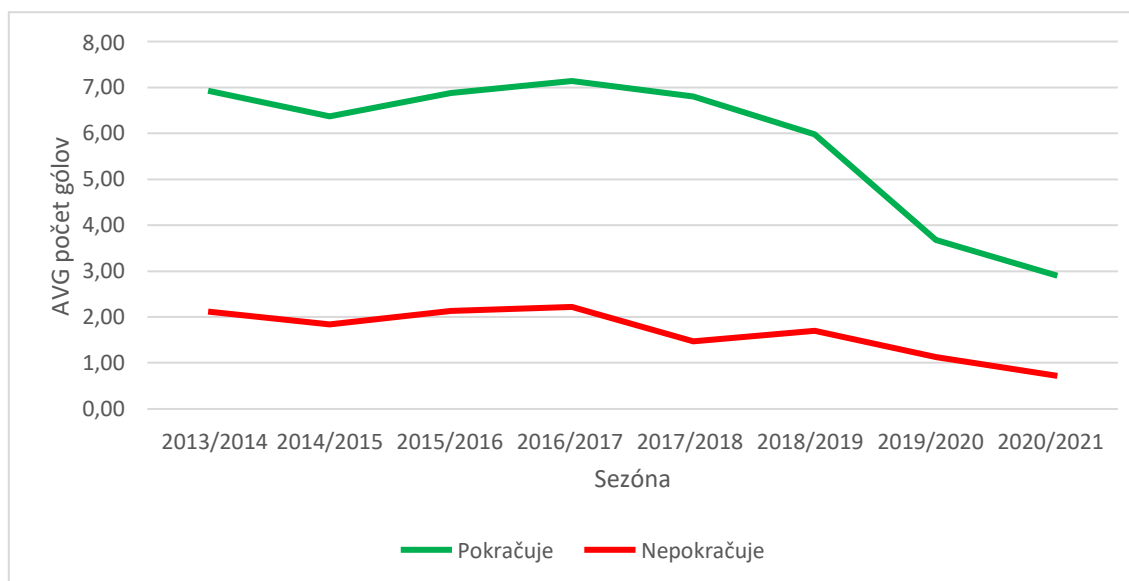
Graf 3.4.3 Priemerný počet nominácií na zápas za sezónu podľa závislej premennej
Zdroj: *vlastné spracovanie na základe zálohy SFZ*

Na Graf 3.4.3 je vidieť **priemerný počet nominácií na zápas**, zoskupený podľa pokračujúcich športovcov a tých, ktorí nepokračujú v organizovanom futbale v SFZ. Graf 3.4.3 ukazuje trend, kde sa **znižuje rozdiel medzi počtom nominácií na zápas mladých futbalistov**, ktorí pokračujú a tých, ktorí nepokračujú. To znamená, že faktor počet nominácií ako determinant pokračovania **znižuje mieru vysvetlenia**. Dôležité je poznamenať, že posledné dve sezóny zobrazené na grafe (rozdiel 2,2 a 1,4 nominácií medzi kategóriami) môžu byť čiastočne ovplyvnené **pandémiou**.



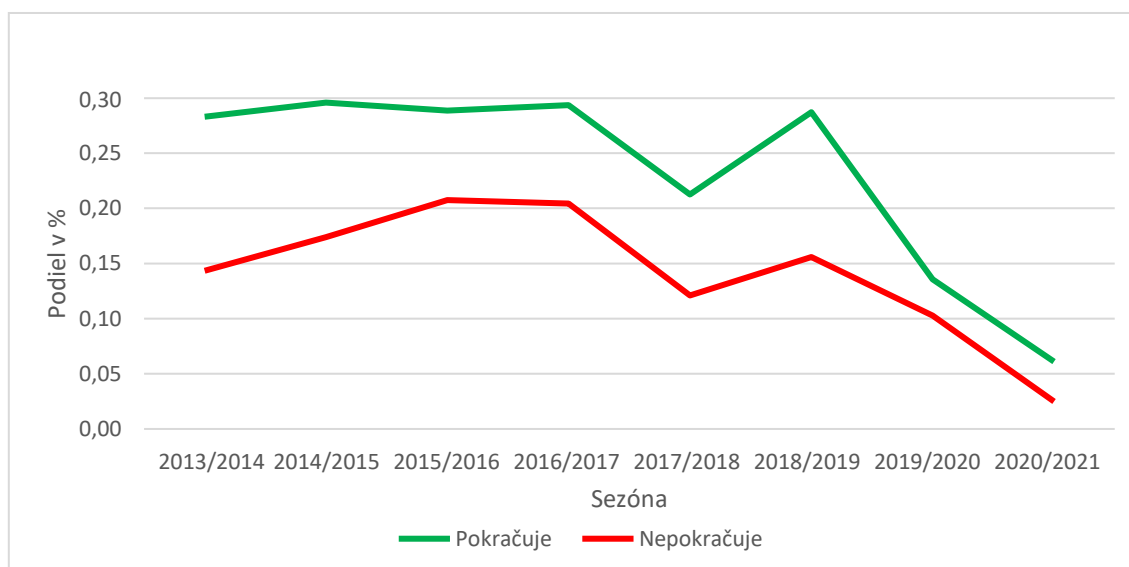
Graf 3.4.4 Priemerný počet zápasov v základnej zostave podľa závislej premennej
Zdroj: *vlastné spracovanie na základe zálohy SFZ*

Graf 3.4.4 ukazuje **priemerný počet štartov hráča v základnej zostave**. Rovnako ako pri Graf 3.4.3, priemerný počet nominácií na zápas môže byť posledné dve sezóny ovplyvnený pandemiou. Grafy majú veľmi podobný vývoj, kde odlišenie medzi nomináciou a základnou zostavou predstavuje **stabilnejší rozdiel** (v intervale od 4,9 po 6,2) medzi skupinou, ktorá pokračuje a nepokračuje. Pri tvorbe modelu bude potrebné zvážiť, ktorý faktor sa použije, pretože faktory **navzájom medzi sebou korelujú**.



Graf 3.4.5 Priemerný počet gólov za sezónu podľa závislej premennej
Zdroj: *vlastné spracovanie na základe zálohy SFZ*

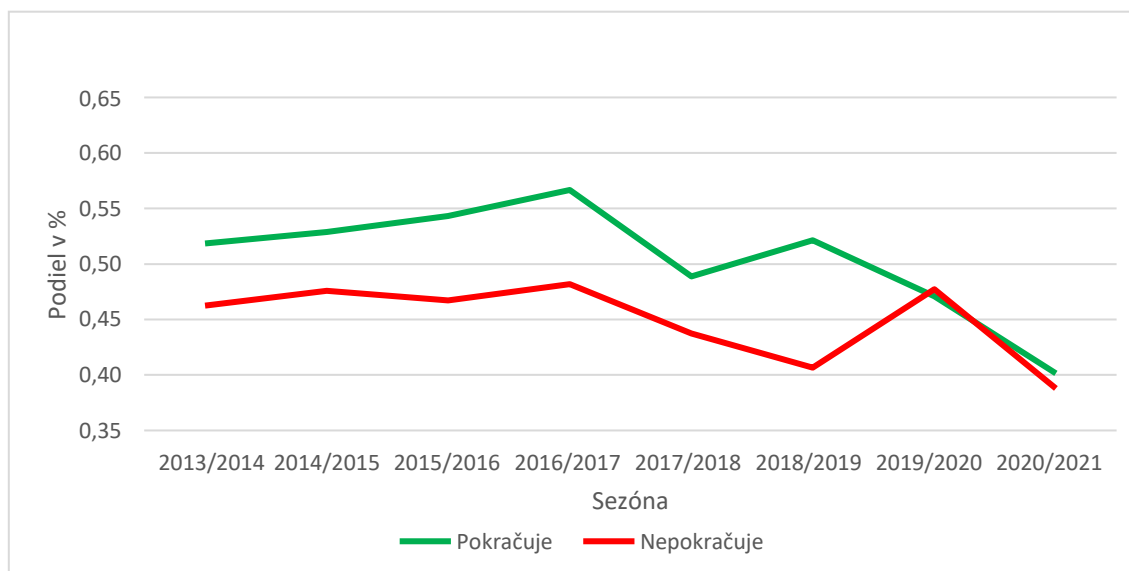
Graf 3.4.5 ukazuje vývoj **priemerného počtu gólov žiakov** (bližšia špecifikácia v časti *Spracovanie dát*) za sezóny od 2013/2014 až po 2020/2021. Pred začiatkom pandémie (sezóna 2019/2020) rozdiel medzi pokračujúcimi a nepokračujúcimi mladými športovcami predstavoval približne **4,5 gólu**. Z toho dôvodu existuje predpoklad vzťahu **nezávislej premennej počet gólov** a pokračovania s organizovaným futbalom.



Graf 3.4.6 % podiel žiakov na viacerých pozíciách za sezóny

Zdroj: *vlastné spracovanie na základe zálohy SFZ*

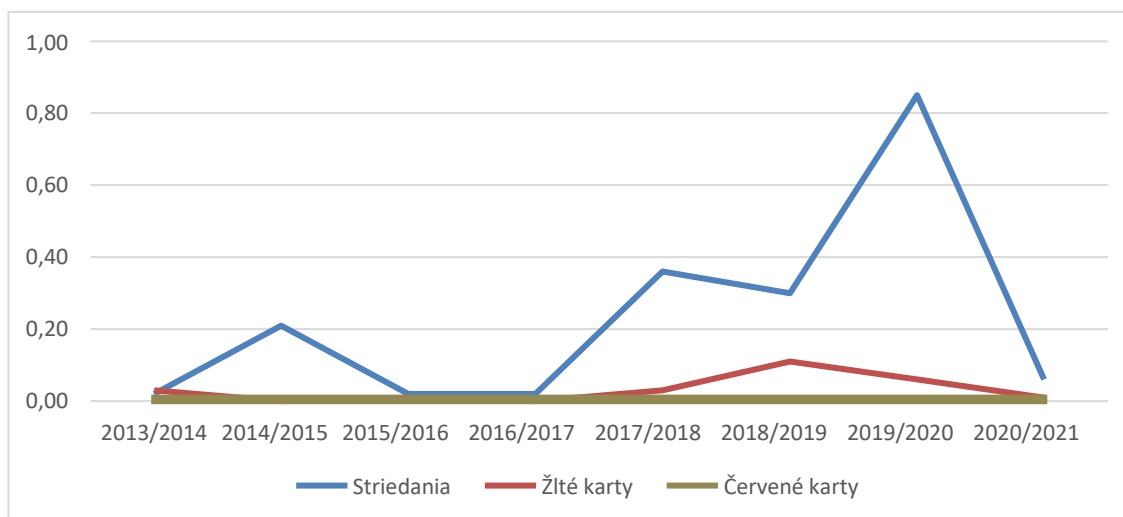
Graf 3.4.6 ukazuje podiel žiakov, ktorí počas jednej sezóny nastúpili na viac ako jednej pozícii, zoskupený podľa pokračovania s organizovaným futbalom. Z faktoru pozície, na ktorej mladý športovec nastupuje počas zápasov, práve sledovanie parametru **viacerých pozícií vykazuje najväčšie odchýlky**, medzi pokračujúcimi (vyššie percento žiakov, ktorí sú počas sezóny na viacerých pozíciách pokračuje) a nepokračujúcimi žiakmi. Rovnako ako pri prechádzajúcich grafoch v období pandémie sa rozdiel medzi pokračujúcimi a nepokračujúcimi žiakmi znižuje a percento žiakov na viacerých pozíciách klesá.



Graf 3.4.7 % podiel žiakov na pozícii záložník za sezónu podľa závislej premennej

Zdroj: *vlastné spracovanie na základe zálohy SFZ*

Pozícia záložník vykazovala najväčšie rozdiely, zobrazená na Graf 3.4.7. Tieto rozdiely sú nižšie (do sezóny 2019/2020 priemerný pozitívny rozdiel na úrovni **7 %**) ako pri atribúte viacerých pozícií (do sezóny 2019/2020 pozitívny rozdiel na úrovni **11 %**) na Graf 3.4.6. **Ostatné pozície vykazovali minimálne rozdiely** medzi kategóriou pokračujúci a nepokračujúci žiaci, útočníci (rovnaké obdobie priemerný rozdiel 3 %) a obrancovia (2 %).



Graf 3.4.8 Zobrazenie rozdielov ostatných nezávislých premenných
Zdroj: *vlastné spracovanie na základe zálohy SFZ*

Zvyšné dostupné štatistické údaje, ktoré sú zapisované v dostupných dátach od SFZ, sú zobrazené na *Graf 3.4.8*. Pre lepšiu výpovednú hodnotu sa rozhodlo zobraziť len **rozdiely medzi priermi skupín pokračujúcich a nepokračujúcich žiakov**. Z *Graf 3.4.8* sa môže vyvodiť záver, že jediný štatistický údaj, ktorý vykazuje vyššie rozdiely, sú **striedania**, kde rozdiel však nie je stabilný a dosahuje aj kladné a záporné hodnoty (raz v priemere striedajú viac pokračujúci a raz nepokračujúci žiaci). Červené karty majú malý výskyt (v priemere 0,01 na hráča) a aj preto je ich výpovedná hodnota minimálna. Žlté karty majú väčší výskyt ako červené (v priemere 0,25 na hráča), ale rovnako ako pri striedaní je hodnota rozdielu priemeru kladná aj záporná.

3.4.1 Zhrnutie dolovania dát SFZ

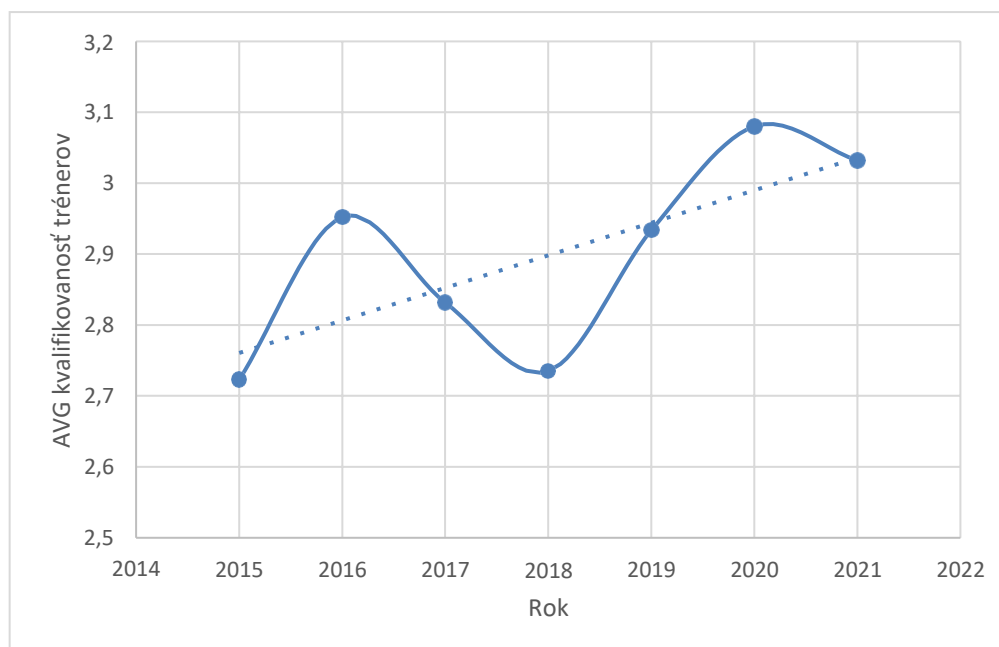
Priemerný vek registrácie športovcov do Slovenského futbalového zväzu **klesá**. Trend pretrváva dlhodobo, kde ročníky 1999 sa registrovali v priemere vo veku 10,84, pričom ročníky 2009 už len 9,37. To znamená, že priemerný vek registrácie športovcov **klesol o 1,47**. Pri sledovaní priemerného veku registrácie **žiacov** sa ukázal rozdiel medzi pokračujúcimi a nepokračujúcimi žiakmi, kde pokračujúci sa v priemere zaregistrovali **o 0,6 roku skôr** ako nepokračujúci. Ďalšie sledované kritériá sú **ovplyvnené pandemiou**, kde rozdiel medzi predchádzajúcimi obdobiami a poslednými dvoma sezónami (2019 - 2021) je citeľný. Priemerný počet nominácií na zápasu priemerný počet zápasov v základnej zostave je rozdielny medzi pokračujúcimi a nepokračujúcimi pred pandemiou na úrovni **5,06** pre nominácie a až **5,37** pre „základ“. **Počas pandémie rozdiel klesol** pri nominácii na 1,80 a pri „základe“ 1,44 v **prospech pokračujúcich**. Priemerný počet gólov žiakov, ktorí pokračujú, je väčší o **4,77** oproti nepokračujúcim pred pandemiou. Počas pandémie sa odohralo menej zápasov, čo predstavuje aj menej príležitostí streliť gól, a tak rozdiel medzi pokračujúcimi a nepokračujúcimi sa znížil na 2,37. Faktor **pozície**, na ktorej žiak nastupuje, vykazuje **minimálne rozdiely**. Jediný post, ktorý pred pandemiou vykazoval vyššie rozdiely v počte (v priemere 7 %) medzi pokračujúcimi a nepokračujúcimi, sú záložníci. Zaujímavejším faktorom je nastupovanie v zostave na **viacerých pozíciách**. Žiaci, ktorí pred pandemiou nastupovali na viac ako jednej pozícii (napr. záložník a útočník), mali **vyššiu pravdepodobnosť pokračovania** oproti tým, ktorí hrali len na jednej pozícii. Ostatné faktory (žltá karta, druhá žltá karta, červená karta, striedania) vykazujú len minimálne rozdiely, a preto ich výpovedná hodnota pre vysvetlenie odchodu žiakov z organizovaného futbalu v SFZ je minimálna.

3.5 Analýza dát z reportov „Šport v číslach“

Centrum vedecko-technických informácií SR poskytuje každoročný report, **šport v číslach**, ktorý je dostupný v období od 2015 po rok 2021. Dáta poskytujú informácie, ktoré **rozširujú základnú databázu dát** získaných zo Slovenského futbalového zväzu. Pre účel práce sa z ročných reportov vybrali len dáta týkajúce sa športu **futbal**.

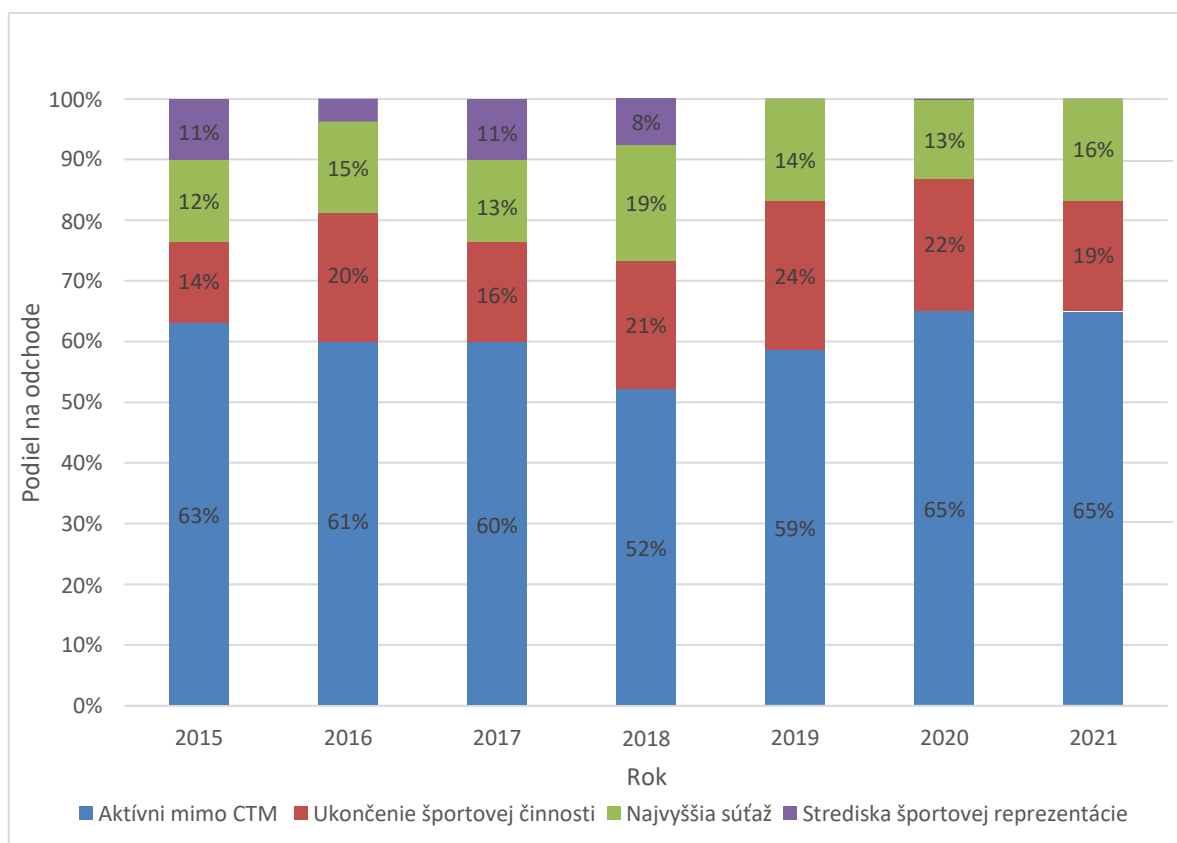
Do útvarov talentovanej mládeže patria centrá olympijskej prípravy, centrá talentovanej mládeže, zväzové centrá talentovanej mládeže, športové triedy, školské športové kluby a podobné. Úlohou týchto útvarov je vytváranie nadštandardných podmienok pre **vzdelávanie a prípravu športových talentov**. Zabezpečenie týchto útvarov pozostáva zo športových zväzov, klubov, zriaďovateľov škôl a Ministerstva školstva, vedy, výskumu a športu SR. (Útvary talentovanej mládeže [on-line])

Zverejnené informácie o počtoch trénerov a odborných zamestnancov sa nachádzajú v reportoch Šport v číslach. Zaujímavým zistením, okrem **konštantného rastu počtu zamestnancov** (okrem roku 2021), predstavuje **vývoj kvalifikácie trénerov**, kde na *Graf 3.5.1* je vidieť **pozitívny trend**.



Graf 3.5.1 Priemerná kvalifikácia trénerov CTM 2015-2021
Zdroj: *vlastné spracovanie na základe reportov Šport v číslach*

Report Šport v číslach zaznamenáva taktiež **dôvod odchodu jednotlivých športovcov** z CTM, kde existujú štyri kategórie dôvodu odchodu: prechod do strediska štátnej športovej reprezentácie, prechod do družstiev najvyššej súťaže, **dalej aktívny (mimo CTM)** a **ukončenie športovej činnosti**. Obzvlášť dôležitú hodnotu pre kontext práce nesú posledné dva dôvody odchodu. Na *Graf 3.5.2* sú zobrazené výsledky dôvodov odchodu športovcov z CTM, kde hodnoty pre strediská štátnej športovej reprezentácie s menším ako 5 % podielom nie sú zobrazené.

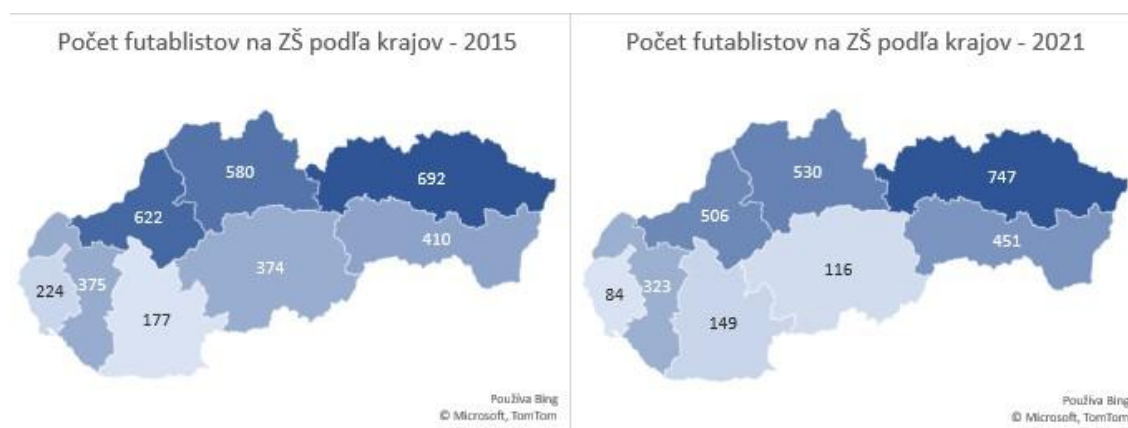


Graf 3.5.2 Kategorizovanie dôvodov odchodu športovcov z CTM 2015-2021

Zdroj: *vlastné spracovanie na základe reportov Šport v číslach*

Výpovedná hodnota *Graf 3.5.2* spočíva najmä v definovaní, **kam športovci po opustení CTM smerujú**. S touto informáciou následne organizácie zastrešujúce CTM pracujú podľa potreby, pretože nie každý dôvod odchodu je negatívny. Napríklad dosahovanie väčšieho podielu dôvodu odchodu - strediska štátnej športovej reprezentácie môže CTM považovať ako **úspech**, na rozdiel od ukončenia športovej činnosti, ktorý je **negatívny**.

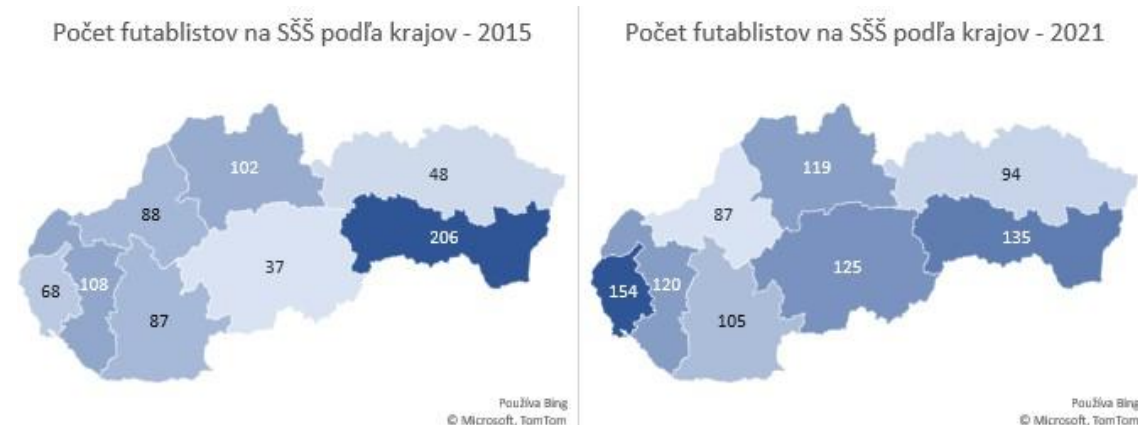
Počet žiakov, ktorí navštevujú futbalovú triedu na základnej škole na Slovensku, sa **znížil o 548**, kde v roku 2015 navštevovalo športové triedy 3454 žiakov oproti roku 2021, kde to bolo už len 2906 žiakov. **Klesajúci trend** pretrváva už od roku 2015, ale k pravidelnosti znižovania došlo až od roku 2018 (približne 5 %). Existuje predpoklad, že zníženie mohlo byť spôsobené okrem iných vplyvov aj **pandémiou**. Nasledujúci report 2022 môže napovedať viac o veľkosti vplyvu faktora pandémie. Na *Obrázok 3.5.1* Počet futbalistov na ZŠ podľa krajov porovnanie roku 2015 a 2021 vidieť, že kraje **východného Slovenska** ako jediné dosiahli **rast** oproti roku 2015.



Obrázok 3.5.1 Počet futbalistov na ZŠ podľa krajov porovnanie roku 2015 a 2021

Zdroj: *vlastné spracovanie na základe reportov Šport v číslach*

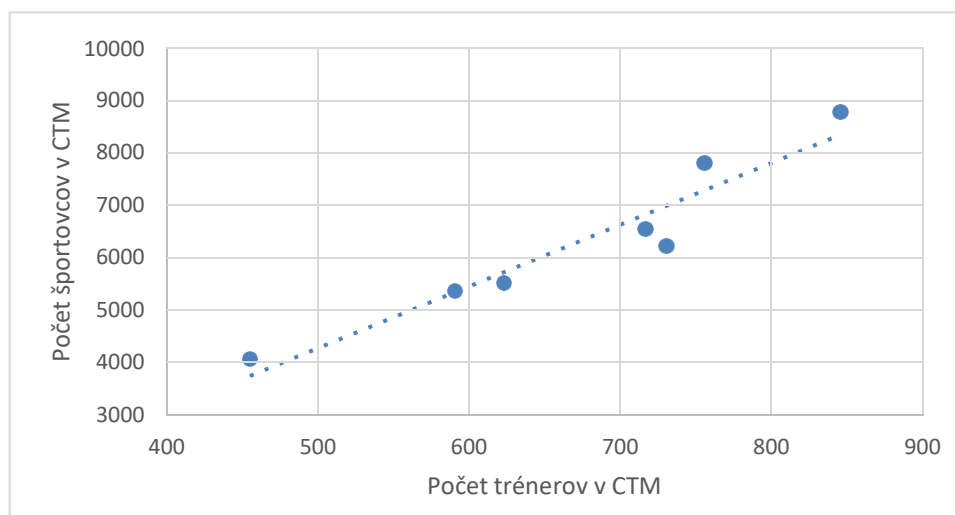
O podobnom trende sa môže hovoriť aj pri stredných školách (SŠ), kde faktorom poklesu okrem pandémie (rok 2018 zaznamenal 0,52 % pokles oproti minulému roku a rok 2019 0,42 %) môže byť aj **zmena zapisovania údajov medzi SŠ a SŠŠ** (strednými športovými školami). Okrem dát podľa krajov toto tvrdenie podporuje aj rozloženie žiakov podľa ročníka, kde počet prvákov na SŠ je každým rokom menší (zvýšenie počtu prvákov sa prejavuje na celkovom počte žiakov na SŠ, kde maximálny efekt je s posunom o 3 roky), pri zachovaní približne rovnakej miery odchodovosti medzi ročníkmi. Na **Obrázok 3.5.2** taktiež vidieť zmenu v období 2015 až 2021, kde distribúcia žiakov na SŠŠ je rovnomernejšia a **počet žiakov na SŠŠ sa všeobecne zvýšil**, kde na západo Slovensku počet **vzrástol o 115 žiakov**, na strednom Slovensku (Žilinský a Banskobystrický kraj) **vzrast o 105 žiakov** a na východnom Slovensku (Prešovský a Košický kraj) **pokles o 25 žiakov**. Dáta SŠŠ ukazujú **opačný trend migrácie** ako ZŠ.



Obrázok 3.5.2 Počet futbalistov na SŠŠ podľa krajov porovnanie roku 2015 a 2021

Zdroj: *vlastné spracovanie na základe reportov Šport v číslach*

Dáta šport v číslach potvrdzujú hypotézu, ktorá sa nedá overiť na úrovni dát zo SFZ. Hypotéza znie: **Počet trénerov v CTM vplýva** (pozitívne koreluje) **na celkový počet športovcov v CTM**. Ako je možné vidieť na **Graf 3.5.3**, naozaj **existuje lineárny vzťah**, ktorý hovorí o tom, že keď rastie počet trénerov, rastie aj počet športovcov a opačne. To znamená, že jeden faktor vplýva na druhý.



Graf 3.5.3 Vplyv počtu trénerov na počet športovcov v CTM
Zdroj: *vlastné spracovanie na základe reportov Šport v číslach*

Na druhej strane hypotéza o vzťahu medzi úrovňou kvalifikácie trénerova odchodom, príchodom (korelácia 0,37) alebo celkovým počtom žiakov (korelácia 0,53) **nemôže byť potvrdená**, preto sa prijíma H_0 : **Vzťah medzi kvalifikáciou trénerov a počtom športovcov neexistuje.**

3.5.1 Zhrnutie analýza reportov Šport v číslach

Analýza dát z reportov šport v číslach za obdobie 2015 - 2021 priniesla zistenia **pozitívneho vplyvu počtu trénerov na počet športovcov**. Zistilo sa, že aj keď tréneri zvyšujú svoju kvalifikovanosť, **nedá potvrdiť hypotéza, že úroveň priemernej kvalifikácie trénerov vplýva na počet športovcov**. Vývoj migrácie športovcov má **opačný trend na ZŠ ako pri SŠŠ**, kde pri základných školách sa počet športovcov znižuje všade okrem východného Slovenska a pri stredných športových školách všade počet športovcov rastie okrem východného Slovenska. CTM využívajú **zaujímavý spôsob klasifikácie odchodu športovcov z CTM**, ktorý by sa mohol aplikovať aj pri SFZ s miernymi úpravami.

3.6 Modelovanie odchodu mládeže z organizovaného futbalu

Pred vytvorením jednotlivých modelov sa použilo **kódovanie jednotlivých premenných**. Tento krok sa použil pre zlepšenie prehľadnosti a zlepšenie interpretácie jednotlivých skriptov. V nasledujúcej *Tabuľka 3.6.1* sa nachádza kód, ktorý predstavuje reprezentáciu premenných v R a názov, ktorý je popísaný bližšie v časti *Spracovanie dát*.

Tabuľka 3.6.1 Kódovanie závislých a nezávislých premenných

Kód	Názov	Kód	Názov
y	Pokracuje	x7	RedCard
x1	old_when_start	x8	Obranca
x2	app_in_nom	x9	Zaloznik
x3	app_in_zak	x10	Utocnik
x4	Substitution	x11	viac_pozicii
x5	Goal	x12	transfer_minulu
x6	YellowCard	x13	norm_narodenie

Zdroj: *vlastné spracovanie*

V *Tabuľka 3.6.2* sú zobrazené základné údaje (funkcia summary) tréningových dát. Dôležitými zisteniami, ktoré z toho vyplynuli, sú: priemer premennej s názvom **y** – **0,73**, čo znamená **nerovnomerné rozdelenie**, kde hráči, ktorí pokračujú v ďalšej sezóne, sú zastúpení oproti hráčom, ktorí nepokračujú, v pomere 3:1. Ďalej rozdiel medzi percentilom 75 a 25 pri premennej **x2** (počet nominácií na zápas) je 6 ale pri premennej **x3** (počet zápasov v základnej zostave) je 13. Premenná **x6** (žlté karty) má nízky priemer a **x7** (červené karty) dokonca len 0,01. Predpoklad nízkeho vysvetlenia.

Tabuľka 3.6.2 Základné štatistické údaje tréningovej množiny

Názov	Počet	Priemer	Štandardná Odchýlka	MIN	Percentil 25	Percentil 75	MAX
y	20134	0,73	0,44	0	0	1	1
x1	20134	8,80	1,60	5	8	10	14
x2	20134	27,00	10,00	1	22	28	104
x3	20134	18,00	8,50	1	11	24	54
x4	20134	1,70	3,00	0	0	2	22
x5	20134	4,90	8,50	0	0	6	127
x6	20134	0,25	0,68	0	0	0	9
x7	20134	0,01	0,08	0	0	0	2
x8	20134	0,43	0,50	0	0	1	1
x9	20134	0,50	0,50	0	0	1	1
x10	20134	0,31	0,46	0	0	1	1
x11	20134	0,23	0,42	0	0	0	1
x12	20134	0,08	0,28	0	0	0	1
x13	20134	-0,36	0,27	-1	-0,6	-0,2	0

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

V *Tabuľka 3.6.3* sú zobrazené základné štatistické údaje pre testovacie dáta.

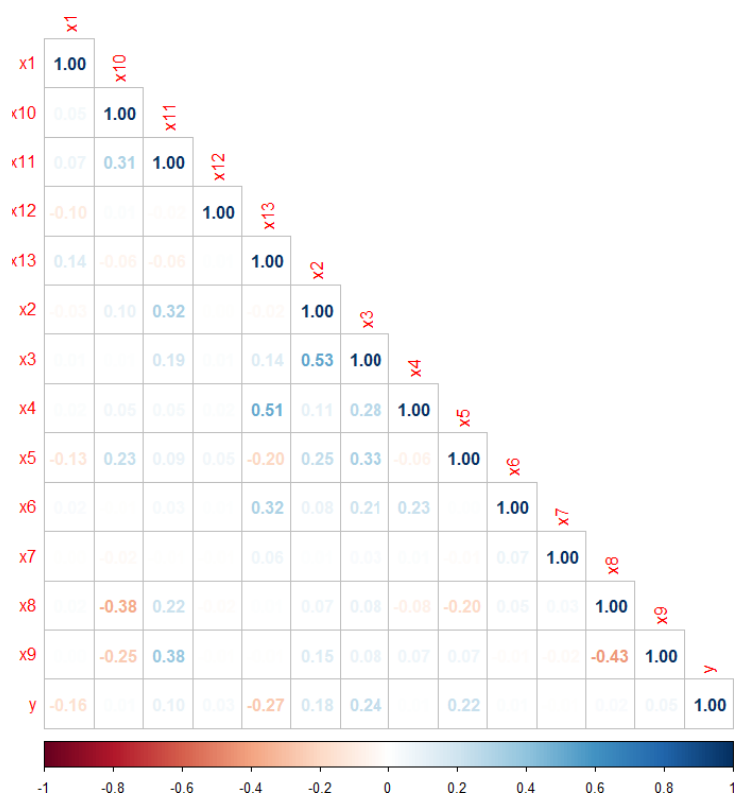
Tabuľka 3.6.3 Základné štatistické údaje testovacej množiny

Názov	Počet	Priemer	Štandardná Odchýlka	MIN	Percentil 25	Percentil 75	MAX
y	2800	0,84	0,36	0	1	1	1
x1	2800	8,30	1,50	5	7	9	14
x2	2800	17,00	5,60	11	15	22	45
x3	2800	5,60	1,70	0	5	7	13
x4	2800	0,63	1,20	0	0	1	6
x5	2800	1,90	3,20	0	0	2	35
x6	2800	0,06	0,27	0	0	0	3
x7	2800	0,01	0,04	0	0	0	1
x8	2800	0,39	0,49	0	0	1	1
x9	2800	0,40	0,49	0	0	1	1
x10	2800	0,27	0,44	0	0	1	1
x11	2800	0,06	0,23	0	0	0	1
x12	2800	0,04	0,18	0	0	0	1
x13	2800	-0,43	0,31	-1	-0,75	-0,25	0

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Rovnako ako pri tréningových dátach je v závislej premennej zastúpená vo väčšej miere (84 %) kategória pokračuje. Premenná **x3** už nemá taký výrazný rozdiel medzi percentilom 75 a 25 (2) ako pri testovacích dátach (13.) Premenné X6, X7 a X12 majú priemer menší ako 0,10. Rovnako aj premenná **X11**, ktorá hovorí o tom, či hráč pôsobil v jednej sezóne na viac ako jednom poste (napr. x8 a x10).

Na Obrázok 3.6.1 je zobrazená **korelačná matica** vygenerovaná v prostredí R. Najdôležitejšie informácie nesie riadok s *premennou* y, kde dochádza ku korelácii medzi y (závislou premennou) a x_n (vysvetľujúcimi premennými). Premenná y koreluje s premennými: **x1**(-0,16), **x2**(0,18), **x3**(0,24), **x5**(0,22), **x11**(0,10) a **x13**(-0,27). Medzi ďalšie dôležité údaje patrí **vzájomná korelácia medzi vysvetľujúcimi premennými**. Premenné x2 a x3 dosiahli medzi sebou najväčšiu koreláciu (0,53). Tieto premenné sa označujú ďalej ako **korelované premenné**.



Obrázok 3.6.1 Korelačná matica tréningovej množiny Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Počet nominácie do zápasu (**x2**) koreluje (0,53) s počtom zápasov v základnej zostave (**x3**), čo sa aj predpokladalo. Počet nominácií (**x2**) koreluje tiež (0,25) s počtom gólov (**x5**) a (0,32) či hráč hral viac ako na jednej pozícii (**x11**). Gól koreluje s väčšinou premenných, kde najvýznamnejšie sú (0,33) počet zápasov v základnej zostave (**x3**), (0,23) pozícia útočník (**x10**) a (-0,20) pozícia obranca (**x8**). Najviac (0,38) na rôznych pozíciách (**x11**) nastupoval záložník (**x9**) a naopak najmenej (-0,43) obranca (**x8**). Počet striedaný (**x4**) pozitívne koreluje (0,51) s normovaným vekom hráča (**x13**). Počet červených kariet (**x7**) nedosiahol zo žiadnou premennou koreláciu väčšiu ako 0,10. Počet žltých kariet (**x6**) koreluje (0,32) s normovaným vekom (**x13**) a (0,23) počtom striedaní (**x4**).

3.6.1 Logistická regresia

Prvý model logistickej regresie **vychádzal zo všetkých premenných**, ktoré sa zistili z dát poskytnutých SFZ na odhalenie štatisticky významných parametrov. Model ukázal, že **parametre x7 – x11 nie sú štatisticky významné** (majú *p-hodnotu* > 0,3). Ostatné vysvetľujúce premenné sú štatisticky významné **minimálne na hladine alfy 0,05**. Na Obrázok 3.6.2 vidieť koeficienty pre jednotlivé premenné.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.207506   0.242842  -0.854   0.3928
x1           -0.157619   0.010977 -14.359 < 2e-16 ***
x2            0.006039   0.002389   2.528  0.0115 *
x3            0.052230   0.002861  18.256 < 2e-16 ***
x4            0.088627   0.007502  11.813 < 2e-16 ***
x5            0.086058   0.005227  16.464 < 2e-16 ***
x6            0.131941   0.027928   4.724 2.31e-06 ***
x7            0.017295   0.195866   0.088  0.9296
x8            0.214440   0.217536   0.986  0.3242
x9            0.097992   0.218453   0.449  0.6537
x10           -0.161535   0.217456  -0.743  0.4576
x11            0.156819   0.228248   0.687  0.4920
x12            0.162432   0.067805   2.396  0.0166 *
x13           -2.976937   0.084542 -35.213 < 2e-16 ***

```

Obrázok 3.6.2 Výstup logistická regresia – koeficienty
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Model dosiahol **AIC skóre 19373**, kde v Tabuľka 3.6.4 sa nachádzajú údaje o celkovej presnosti, senzitivite a špecificite modelu pri **rôznych úrovniach prahovej hodnoty** na tréningových a testovacích dátach.

Tabuľka 3.6.4 Výstup logistického modelu - všetky premenné

Množina – prahová hodnota	Celková	Senzitivita	Špecificita
Train - 0,4	77,02	97,02	23,07
Train - 0,5	77,85	93,28	36,24
Train - 0,6	76,98	86,47	51,40
Test - 0,4	74,76	80,25	45,59
Test - 0,5	65,14	66,98	55,33
Test - 0,6	56,29	54,52	65,76

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Ďalší model využíval len **štatisticky významné premenné** z predchádzajúceho modelu. Tento model dosiahol **AIC skóre 19443**. V Tabuľka 3.6.5 sa nachádzajú hodnoty sledovaných metrík pre jednotlivé prahové hodnoty na tréningových a testovacích dátach.

Tabuľka 3.6.5 Výstup logistického modelu - štatisticky významné premenné

Množina – prahová hodnota	Celková	Senzitivita	Špecificita
Train - 0,4	76,95	97,06	22,69
Train - 0,5	77,72	93,34	35,58
Train - 0,6	76,81	86,42	50,90
Test - 0,4	74,50	80,00	45,12
Test - 0,5	64,43	66,34	54,20
Test - 0,6	56,60	54,98	65,31

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Posledný vytvorený logistický model, zobrazený v *Tabuľka 3.6.6*, je zložený len z **korelovaných premenných**, ktoré dosiahli pozitívnu alebo negatívnu koreláciu minimálne na úrovni 0,16 a vyššiu (viď *Obrázok 3.6.1*).

Tabuľka 3.6.6 Výstup logistického modelu - korelované premenné

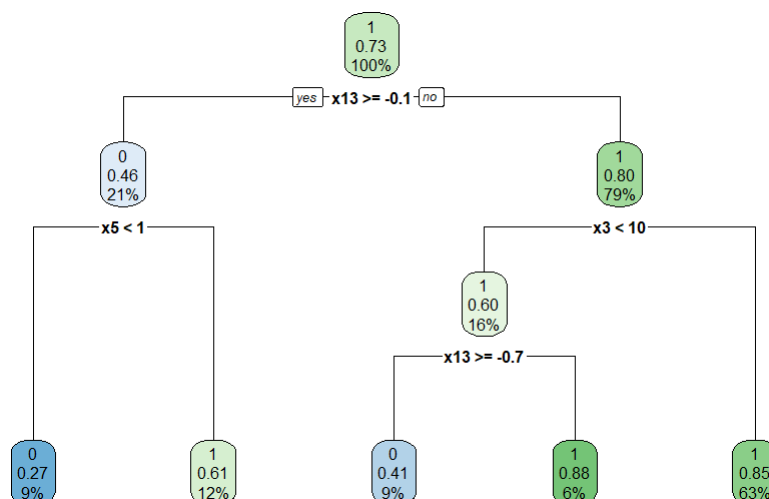
Množina – prahová hodnota	Celková	Senzitivita	Špecificita
Train - 0,4	76,76	97,54	20,67
Train - 0,5	77,33	93,41	33,93
Train - 0,6	76,85	86,42	51,01
Test - 0,4	75,36	81,14	44,44
Test - 0,5	64,57	66,43	54,65
Test - 0,6	55,50	53,63	65,53

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Model dosiahol najvyššiu hodnotu **AIC skóre 19582**, kde rozdiely v celkovej miere vysvetlenia, senzitivite a špecificite sú minimálne.

3.6.2 Rozhodovací strom

Pri vytváraní modelu založenom na klasifikácii podľa rozhodovacieho stromu sa najskôr použili **všetky dostupné premenné**. Následne sa pomocou metrík celkovej presnosti, špecificity a senzitivity modelu porovnali modeli vytvorené len so štatisticky významnými premennými z logistickej regresie a **korelovanými premennými**. Všetky modely dosiahli na testovacích dátach podobné výsledky v sledovaných metrikách. Z toho dôvodu sa vybral model s najmenším počtom vysvetľujúcich premenných. Na *Obrázok 3.6.3* je možné vidieť jednotlivé **rozdelenia** (splits) v rozhodovacom strome.



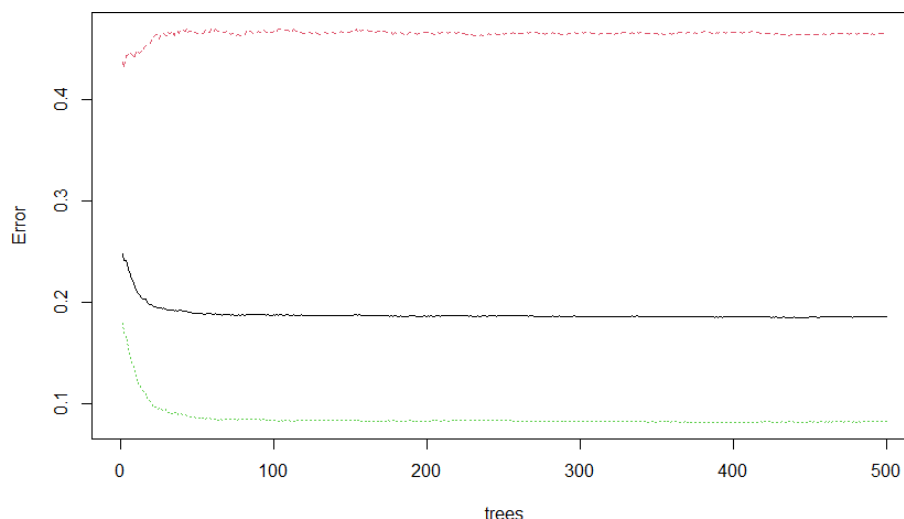
Obrázok 3.6.3 Výsledný rozhodovací strom

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Pri optimalizovaní na zlepšenie modelu sa použil vo **funkcii rpar** aj parameter control, ktorý hovorí o **minimálnom počte splitov**, **maximálnej hĺbke stromu** alebo **minimálnom počte pozorovaní v liste poslednej úrovne**. Ani jeden z atribútov parametru control však nepomohol vytvoriť lepší model. Finálny model kategorizácie podľa rozhodovacieho stromu tak dosiahol metriky na úrovni **celkovej presnosti** modelu **71,61**, **špecificite 55,33** a **senzitivite 74,65**.

3.6.3 Náhodné lesy

Náhodné lesy vo väčšine prípadov dosahujú lepšie výsledky ako rozhodovacie stromy, z toho dôvodu sú použité ako ďalší druh klasifikácie. Pred vytvorením modelu náhodných lesov sa musela **konvertovať závislá premenná na faktor**. Na Obrázok 3.6.4 je zobrazený vývoj chýb (y) v jednotlivých stromoch (x).



Obrázok 3.6.4 Vývoj chýb počas vytvárania jednotlivých stromov
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Vytvorený model náhodných lesov dosiahol na testovacích dátach nasledujúce štatistiky: **kappa** na úrovni **0,2139**; **celková presnosť modelu 0,8046**; **špecifická 0,3016**; **senzitivita 0,8987** a **vyváženosť modelu 0,6001**.

3.6.4 Metóda podporných vektorov

Ako posledná metóda klasifikácie je použitá SVM (metóda podporných vektorov). Nakoľko sa jedná o dáta, ktoré sa **nedajú rozdeliť lineárne** (kategorická premenná), využil sa **kernel** druhu radial a polynomial. Pre optimalizovanie modelov SVM sa taktiež využil **tunning** (funkcia `svm_tune`), kde parametre **gamma** (ako ďaleko dosiahne vplyv jedného tréningového príkladu) a **cost** (kompenzácia správnej klasifikácie na tréningových dátach a maximalizácia nákladov rozhodovacej funkcie) sa dosadili hodnoty do jednotlivých druhov kernel triku pre oba parametre z intervalu (0,001; 0,01; 0,1; 1; 5; 10). Metrikami, ktoré sa najviac sledovali počas hodnotenia modelu, sú: **špecifická** z dôvodu nevyváženosti kategórií a tiež z dôvodu zamerania sa na odhalenie hráčov, ktorí nepokračujú ďalej s organizovaným futbalom v SFZ. **Kappu**, ktorá je pri modeloch, ktoré sú nevyvážené, významným ukazovateľom. Vyváženosť kategorizácie (balance accuracy) a celková presnosť modelu, ktoré majú výpovednú hodnotu len v kombinácii so **špecificitou**.

V *Tabuľka 3.6.7* sa nachádzajú ukážky tuningu jednotlivých parametrov **gamma** a **cost**.

Tabuľka 3.6.7 Výsledky tuningu SVM modelu - kernel trik radial

Radial					
Gamma_Cost	Celková	Špecificita	Senzitivita	Kappa	Bal Acc
01_001	0,765	0,362	0,840	0,187	0,601
01_01	0,688	0,507	0,721	0,163	0,614
01_1	0,662	0,612	0,672	0,183	0,642
01_5	0,648	0,658	0,646	0,186	0,651
01_10	0,730	0,522	0,770	0,224	0,646
1_001	0,789	0,286	0,882	0,174	0,584
1_01	0,632	0,626	0,633	0,157	0,629
1_1	0,629	0,646	0,625	0,163	0,636
1_5	0,640	0,635	0,641	0,169	0,638
5_001	0,842	0,000	1,000	0,000	0,500
5_01	0,760	0,392	0,828	0,196	0,610
5_1	0,666	0,596	0,679	0,180	0,630
5_5	0,650	0,607	0,658	0,168	0,632

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Na základe vyššie špecifikovaných sledovacích metrík najlepší SVM model, ktorý využíva radial kernel je ten, ktorý má parametre **gamma 0,1** a **cost 5**. Model je vyvážený v predikovaní oboch kategórií, kde sú dosahované len minimálne rozdiely medzi vyváženosťou (0,651) a celkovou presnosťou modelu (0,648).

Tabuľka 3.6.8 Výsledky tuningu SVM modelu - kernel trik polynomial

Polynomial					
Gamma_Cost	Celková	Špecificita	Senzitivita	Kappa	Bal Acc
01_001	0,838	0,152	0,967	0,163	0,559
01_01	0,768	0,401	0,837	0,214	0,619
01_1	0,741	0,500	0,787	0,227	0,643
01_5	0,735	0,514	0,776	0,226	0,646
1_001	0,731	0,517	0,771	0,221	0,644
1_01	0,732	0,526	0,770	0,227	0,648
1_1	0,725	0,533	0,761	0,221	0,647
1_5	0,730	0,528	0,768	0,226	0,648
5_001	0,726	0,528	0,763	0,221	0,646
5_01	0,696	0,567	0,720	0,201	0,644

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

SVM model, ktorý využíval polynomial kernel, výsledky zobrazené v *Tabuľka 3.6.8*, je najlepší pri parametroch **gamma 1** a **cost 0,1** alebo **5**. Model však dosahuje **horšie výsledky ako radial kernel**, z pohľadu rozdielu medzi špecificitou (0,526) a senzitivitou (0,770) modelu.

3.6.5 Zhrnutie modelovania odchodu mládeže z organizovaného futbalu

Dáta, ktoré sa získali s databázy, sú nevyvážené, čo sa týka závislej premennej (pokračovanie). Z toho dôvodu sa bude pri vytvorených modelov zameriavať na sledovanie **špecifity**. Nebudú sa využívať metódy na **bilancovanie dát** (vynechanie časti množiny hráčov, ktorá pokračuje alebo umelé vytváranie dát pre nepokračujúcich). **Pokračovanie koreluje** s vekom registrácie do organizácie SFZ (-0,16), s normovaným narodením (-0,27), s počtom nominácií na zápas (0,18), s počtom zápasov v základnej zostave (0,24) a počtom gólov (0,22). Využívanie počtu nominácií a zápasov v základnej zostave spoločne ako vysvetľujúce premenné môže spôsobiť problém, pretože spolu korelujú na úrovni 0,53. Tento predpoklad sa však nepotvrdil a pri vytváraní modelov sa používali obe premenné ako vysvetľujúce.

Prvý model logistickej regresie ukázal, že až 8 premenných z 13 sú štatisticky významné na hladine významnosti 0,05 a menej. Pri zjednodušovaní modelu logistickej regresie sa ukázalo, že stačí, aby sa v modeli použili len **korelované premenné** s y na dosiahnutie výsledku len minimálne odlišného od použitia všetkých premenných. Posledný model logistickej regresie dosiahol najlepšie výsledky pri **prahovej hodnote 0,4** (threshold value), kde na testovacích dátach dosiahol AIC skóre 19582, **celkovú presnosť modelu 75,36 %**, senzitivitu 81,14 % a **špecifitu 44,44 %**. Vytváranie rozhodovacích stromov prebiehalo podobne ako pri logistickej regresii. Finálny model rozhodovacích stromov dosiahol výsledky **celkovej presnosti modelu 71,61 %**, senzivitve 74,65 % a **špecifite 55,33 %**. **Náhodné lesy** vo väčšine prípadov poskytujú lepší výsledok ako rozhodovacie stromy. Finálny model rozhodovacieho stromu dosiahol **celkovú presnosť 80,46 %**, senzitivitu 89,87 % a **špecifitu 30,16 %**. Ako posledné sa vytvárali modely metódou podporných vektorov. SVM vytvorila celkovo najlepší model, ktorý vznikol počas modelovania. Pomocou **kernelu radial** sa vytvoril models parametrami gamma 0,1 a cost 5, kde výsledky na testovacích dátach sú na úrovni **celkovej presnosti modelu 64,80 %**, senzivitve 64,60 % a **špecifite 65,80 %**.

Počas modelovania sa využili aj **optimalizačné postupy** pre jednotlivé modely. Každý model dostal okrem surových dát aj **škálované dáta**. Pri SVM sa využil **tunning parametrov** gamma a cost, pri rozhodovacích stromoch sa experimentovalo s parametrom control, ktorý určuje **počet rozdelení** alebo aj **počet dát v poslednom liste**. V zhrnutí sa nachádzajú len modely, ktoré sú podľa stanovených metrík najlepšie pre jednotlivé druhy klasifikácie.

4 NÁVRHOVÁ ČASŤ

Návrhová časť diplomovej práce pozostáva z dvoch hlavných návrhov. Prvým návrhom je získanie potrebných dát pre vysvetlenie odchodu mládeže z organizovaného futbalu v SFZ. Druhým návrhom je zvýšenie počtu dimenzií výstupu, čo vedie k lepšiemu charakterizovaniu odchodu. Každý návrh je zložený z definície, následného popísania atribútov a vytvorenia scenárov.

4.1 Potrebné dáta na vysvetlenie problému

V časti práce *spracovanie dát* sa venujeme spôsobu získavania informácií, vedomostí a znalostí z dát (laženie dát). Táto časť práce ukazuje, ktoré dáta sú dostupné na vysvetlenie odchodu žiakov z organizovaného futbalu v *Slovenskom futbalovom zväze*. Výsledkom analýzy dát, ktorá je detailne vysvetlená v *analytickej časti práce* sa došlo k záveru, že aktuálne dáta dostupné SFZ nie sú postačujúce na vysvetlenie fenoménu odchodu športovcov po dovŕšení 14 – 15. roku veku narodenia z organizovaného futbalu v SFZ. Preto prvá polovica návrhovej časti sa zaoberá možnými spôsobmi získania potrebných dát na dodatočné vysvetlenie odchodu a druhá časť poskytuje návrhy na zlepšenie stavu zisteného z analýzy.

Časť práce *Faktory vplyvajúce na odchod z organizovaného športu* detailne opisuje, ktoré faktory vplyvávajú na rozhodnutie mladých športovcov v pokračovaní alebo skončení s organizovaným futbalom. V tejto časti práce sa **zadefinujú oblasti**, v ktorých je potrebné získať ďalšie dáta pre vysvetlenie problému a **špecifikujú sa zisťované parametre**. Najvhodnejší formát získavania potrebných dát od vybraných športovcov je formou **dotazníkového prieskumu**.

Viacere štatistiky a výskumy ukázali, že u športovcov je dôležitejšie **subjektívne vnímanie faktorov** ako ich objektívne hodnotenie. Z toho dôvodu by sa dáta mali viac zameriavať na **subjektívne vnímanie**, ktoré je škálovo ohraničené v intervale od napr. od 1 po 10 s jednobodovými alebo dvojbodovými rozstupmi. Na začiatku je potrebné respondenta identifikovať. Následne sa získavajú **vstupné údaje** (svoje charakteristiky), medzi ktoré patrí rok a mesiac narodenia, v akom kraji pôsobil, ako dlho hrá/hral futbal, kedy začal s futbalom, najvyššie dosiahnuté vzdelanie a povolanie. Charakteristiky sú vybrané na základe výskytu v štúdiách (VAN YPEREN, N. W. a kol. 2022), (INEKE, D. a kol. 2018) alebo zistenia dôležitosti pri **processe ťaženia dát** alebo **modelovania**.

Po uvedení vstupných charakteristík sa prejde k jednotlivým oblastiam, ktoré sa zistili počas analýzy faktorov vplyvajúcich na odchod z organizovaného futbalu, že predstavujú **dôležité faktory**. Jednotlivé oblasti by neboli od seba oddelené do sekcií, ale nasledovali by logicky po sebe. Prvým definovaným faktorom je **záujem o šport**. Otázky pre faktor by mali byť zamerané hlavne na **vnímanie iných možností mimo futbalu**, **vnímaní pocitu** radosti, šťastia a zábavy počas vykonávania futbalu, **pozitívne** alebo **negatívne vnímanie subjektov v kontexte futbalu** (tréneri, klub, organizácia, rodičia). Ďalšími faktormi sú **náročnosť súťaže** a **trénovanie**. Otázky pre tieto faktory by mali smerovať do oblasti ako do akej miery je/bol futbal vnímaný ako práca, ohodnotenie vlastnej fyzickej vyspelosti v mládežníckych kategóriách oproti rovesníkom, vnímaná vlastná **kvalita oproti rovesníkom počas mládežníckych kategórií**, vnímanie tímového úspechu, vnímanie kvality trénerov a tréningového procesu, vnímanie frekvencie a intenzity tréningov, v akom veku vnímajú začiatok špecializovania na futbal, ak špecializáciu vnímajú. **Ostatné zistené faktory**, medzi ktoré patrí napríklad **pocit vyhorenia** alebo **prerozdelenie času** sa dajú zistiť otázkami, ako sú: vnímaný tlak zo subjektov v kontexte futbalu, očakávania v mládežníckych kategóriách od futbalu (škála by predstavovala najmenej = trávenie času s kamarátmi a najviac = profesionálna kariéra), porovnanie orientácie na školu a futbal (najmenej = škola a najviac = futbal),

počet zranení a ich závažnosť, vnímaná rovnováha medzi futbalom a životom, vnímanie dostávaných príležitostí vo futbale a vnímaná kvalita zariadení a vybavenia.

Zodpovedaním otázok v jadre dotazníka sa získajú odpovede na **vysvetľujúce premenné**. Nakoľko odporúčanie vychádza z modelovania, ktoré je založené na učení s učiteľom, je potrebné zistiť aj vysvetľujúcu (závislú) premennú. Otázkou na získanie tejto informácie je priame dopytovanie respondenta: pokračujete s organizovaným futbalom v SFZ, kde v prípade ak odpoveď je nie nasledujú ďalšie otázky na zistenie dôvodov prečo. Ďalšie otázky v prípade nepokračovania sú v akom veku respondent skončil, vyberte dôvody prečo z možností (definované faktory z analytickej časti plus možnosť iné), scenáre po skončení (výber z možností: prestup do iného zväzu, pokračovanie už len v neorganizovanom futbale, skončenie s vykonávaním futbalu), iný druh angažovania sa vo futbale (tréner, funkcionár a pod.) a v akej miere sleduje futbal.

Detailnejšie zobrazenie otázok, druhu odpovedí a poznámok k otázkam sa nachádza v **prílohe A**. Jednotlivé otázky boli testované na zrozumiteľnosť a chápanie medzi 10 bývalými a aktuálnymi futbalistami pôsobiacimi v SFZ. Spätné väzby z ich strany boli zapracované vo finálnej verzii, ktorá je zobrazená v tabuľke v **prílohe A**.

4.1.1 Spôsob získavania dát

Poskytnuté dáta od SFZ obsahujú viac ako 10.000 športovcov, ktorí sa zúčastnili od sezóny 2012/2013 do sezóny 2021/2022 v súťažiach s názvom, ktorý obsahoval „LMŽ“ alebo „LSŽ“. Niektorí športovci pokračovali, niektorí prestúpili do iných líg, súťaží, niektorí sú stále nejakým spôsobom angažovaní v organizovanom futbale a niektorí skončili s futbalom úplne. SFZ má stále dostupné údaje o všetkých týchto športovcoch, akými sú aj záznamy o nimi odohratých zápasoch, ale aj **PII** (osobné identifikačného údaje) a kontakt. To znamená, že existuje kontakt skoro na celý základný súbor športovcov. SFZ má 3 možnosti, ako získať doplňujúce informácie na vysvetlenie problému. Každá z týchto možností sa bližšie rozoberie v nasledujúcom texte.

Email - online dotazníkový prieskum

Realizácia dotazníkového prieskumu prostredníctvom emailu patrí medzi jednu z najviac využívaných metód zberu dát. Tento spôsob je preferovaný aj z dôvodu, že sa môžu uviesť bližšie detaily, prečo je odpoveď respondenta dôležitá a mal by si nájsť čas na vyplnenie a informovať ho viac o tom, prečo sú tieto dáta dôležité (aký problém riešia). Odporúčaním je využitie **online dotazníku**, ktorý presmeruje respondenta na inú stránku, nie **inline dotazník**, ktorý sa vyplňa priamo v maile. Priemerná miera odpovedí na online prieskumy v oblasti vzdelávania je na úrovni 44 %. (MENG-JIA, W. a kol. 2022)

Medzi **výhody** zasielania online dotazníkov cez email patria: nízke požiadavky na zdroje pri realizácii (nákladová efektívnosť), dotazníky majú menšiu mieru chybovosti oproti tradičným metódam z dôvodu nezasahovania sprostredkovateľa dotazníku, možnosť výberu veľkosti prieskumu (môže sa vykonať aj menší alebo väčší prieskum na zistenie v závislosti od dôležitosti a času na získanie odpovedí). Emailové zasielanie dotazníkov umožňuje využívanie nastavenia plánu automatického posielania, čo je veľká výhoda v prípade pravidelných dotazníkov. (ERHART, M. a kol. 2009)

Najväčšie **nevýhody** sú spoľahlivosť a kvalita údajov. Niektorí respondenti môžu odpovedať na dotazník len za účelom dosiahnutia stimulu. Obmedzený dosah, hlavne pri prvotnom zasielaní, kde sú aj športovci zo sezóny 2012/2013, ktorí už nemusia mať aktuálne emailové adresy alebo športovec emailovú adresu nepoužíva. (ERHART, M. a kol. 2009)

Telefonický rozhovor – dotazníkový prieskum

Získavanie dát prostredníctvom telefonického rozhovoru patrí rovnako medzi metódy zberu dát. V kontexte riešenej problematiky využitie metódy môže poskytnúť

zvýšenie počtu odpovedí po fáze zasielania emailov s online dotazníkom alebo sa môže vykonať samostatne. Telefonický rozhovor pokryje špecifický segment športovcov, ktorí nevyužívajú email, neodpovedajú na email, alebo menili emailovú adresu, ale telefónne číslo majú rovnaké. Existuje predpoklad, že telefónne číslo sa mení menej často ako email. Taktiež respondenti lepšie reagujú a interagujú prostredníctvom hovoru ako mailu. Aj keď niektoré faktory môžu viesť k potrebe zvýšenia miery odpovedí, výskumníci by mali zvážiť, či sa oplatí usilovať sa o vyššiu mieru odpovedí. (MENG-JIA, W. a kol. 2022)

Medzi **výhody** patrí "jedinečnosť", kde ľudia dostávajú denne v priemere stovky emailov, ale len pár telefonátov. Oslovovanie prostredníctvom telefónu vytvára priestor pre väčšiu mieru odozvy ako poslania správy do preplnenej mailovej schránky. Detailnosť a dimenzionalnosť odpovede, vďaka rozhovoru môže sprostredkovateľ získať dotlačne informácie, akými sú tón hlasu, obsiahnejšiu odpoveď, prípadne dovysvetlenie otázky v prípade nepochopenia. (ERHART, M. a kol. 2009)

Nevýhodami sú vyššie náklady na realizáciu, kde je potrebné buď využiť interné zdroje alebo externé služby na zrealizovanie. Sprostredkovateľa hovoru môže navádzať respondent, kde niektoré otázky vyžadujú vlastnú interpretáciu respondenta, aby sa zistilo jeho vnímanie. Nie vysvetlenie vnímania dopytujúceho a následná odpoveď respondenta.

4.1.2 Scenáre návrhu získania dát

Cieľom zberu nových potrebných dát je získať **minimálne vzorku 400 odpovedí**, čo predstavuje približne **5 % chybovosť** pri základnom súbore 10000. Na dosiahnutie tohto cieľa podľa autora stačí využitie dvoch metód zberu dát pre dotazníkový prieskum, a to zasielaním emailov a/alebo formou telefonického dopytovania.

Tabuľka 4.1.1 Sumarizácia scenárov získavania dát

Variant scenáru	Forma získavania dát	Poskytnutie stimulu	Pravdepodobnosť naplnenia cieľu
1	email	nie	70 %
2	email	áno	80 %
3	email + telefón	nie	100 %
4	telefón	nie	100 %

Zdroj: *vlastné spracovanie*

Tabuľka 4.1.1. zobrazuje základné parametre jednotlivých scenárov návrhu spôsobu získania dát. Forma získavania dát predstavuje využitý nástroj na získavanie. Poskytnutie stimulu hovorí o tom, či scenár pracuje s fyzickou formou motivácie respondenta k vyplneniu dotazníka. Pravdepodobnosť vyjadruje percentuálnu mieru splnenia stanoveného cieľu na základe názoru autora a kontextu uvedeného nižšie.

Pre maximalizovanie miery odpovedí zasielaním dotazníkov cez email je vhodné **oznámiť** vybranej vzorke, že dostanú email, vysvetlí im, **prečo sú práve oni vybraní a prečo majú dotazník vyplniť**. Následne po určitom čase od zaslania tí, ktorí dotazník nevyplnili, by mali dostať **pripomienku na vyplnenie**. Otázky a odpovede by mali byť jasné, bez gramatických chýb a **nie príliš náročné na pochopenie**. Faktor, ktorý je pozitívne vplyva na vyplnenia dotazníka je **predpoklad záujmu o problematiku u športovcov, resp. bývalých športovcov**. (SALEH, A. et al. 2017)

V prípade, keď účastníci dostanú **malý darček pred vyplnením dotazníka**, sa zvyšuje šanca vyplnenia dotazníka. Toto tvrdenie je podporené štúdiou, ktorá hovorí o vplyve darčeka na respondentov **pocit povinnej reciprocity**, čo vedie k vyplneniu

dotazníka. Príkladom malého darčeka môže byť: podpísaná fotka futbalistu, lopta, rôzne formy zľavy alebo aj lístky na zápas. (AJILORE, O. H. et al. 2021)

V prípade scenárov 3 a 4 je pravdepodobnosť dosiahnutia požadovaného výsledku 100 %. Dôvodom tohto tvrdenia je primárne zistenie štúdie, kde sa realizoval prieskum miery odpovedí na pacientoch. Výsledkami sú zistenia, kde miera odpovedí pre jednotlivé formy dotazníkového prieskumu je nasledovná: 34 % mail, **24 % email** a až **80 % miera odpovedí na prieskum cez telefón**. (SJOERD, P. F. T. N. 2014) Kombinovaný spôsob získavania dát v 3. scenári by sa realizoval prostredníctvom dvoch fáz. Prvou je získavanie dát email formou, kde chýbajúce odpovede sa získajú formou telefonického dopytovania.

4.2 Zvýšenie dimenzií odchodu zo SFZ

Na odchod zo zväzu sa dá pozerať z rôznych druhov perspektív. Jedna z týchto perspektív sa zaoberá aj ďalšou aktivitou v športe, kde pokračovanie, respektíve nepokračovanie s organizovaným futbalom v Slovenskom futbalom zväze sa rozširuje o **d ďalšie dimenzie**, ktoré sú znázornené na *Obrázok 4.2.1*.



Obrázok 4.2.1 Dimenzie odchodu z organizovaného futbalu
Zdroj: *vlastné spracovanie*

Tento návrh vychádza zo sledovania odchodu z CTM, znázornený na *Graf 3.5.2*, kde je ukázaný v čase podiel jednotlivých dôvodov na odchod športovcov (futbalistov) z CTM. Každá z možností odchodu v CTM sa dá ilustrovať s menšími úpravami aj pre návrh **zvýšenia dimenzií závislej premennej**. V *Tabuľka 4.2.1* sa nachádza navrhované kategorizovanie výstupu.

Tabuľka 4.2.1 Mapovanie dôvodu odchodu z CTM na návrh

CTM	Návrh
Najvyššia súťaž	Vyššia veková kategória (dorast)
Strediská športovej reprezentácie	Prestup do iného zväzu
Aktívny mimo CTM	Zmena športu
Ukončenie športovej činnosti	Skončenie so športom

Zdroj: *vlastné spracovanie na základe reportov „Šport v číslach“*

Po získaní ďalších dát, popísaných v časti práce 4.2, je potrebné vyskúšať kategorizovanie pred aj po predikovaní, ktorí športovci pokračujú. To znamená, že

v **scenári po** je výstup len pokračuje alebo nepokračuje a následne sa pracuje len s tými, ktorí nepokračujú a tí sa zoskupujú. A v **scenári pred** je výstup už rozdelený do jednotlivých kategórií s použitím viacerých dimenzií výstupu.

Návrh zvýšenia počtu dimenzií odchodu zo SFZ môže byť pri scenároch 2 a 3 využívaný aj inými subjektami. Zistením migrácie futbalistov medzi jednotlivými futbalovými zväzmi, poskytuje prehľad na základe, ktoré si zväzy môžu prerozdeľovať zdroje medzi sebou. Nejedná sa len o finančné zdroje, ale taktiež personálne. V prípade nárastu záujmu o futbal v jednom zväze, druhý zväz, ktorý zaznamenal pokles môže poskytnúť chýbajúci personál alebo vybavenie. Dôjde tak k zlepšeniu využívania zdrojov. V scenári 3 výsledky implementácie návrhu môžu využívať nielen futbalové zväzy, ale taktiež zväzy iných športov, ktoré zistia koľko futbalistov prechádza k nim a tak môžu napríklad upraviť tréningový proces alebo prípravu športovcov. Výsledok návrhu môže pomôcť taktiež aj pri prerozdeľovaní finančných prostriedkov medzi uznané športy, kde MINEDU môže slúžiť ako ďalší parameter vzorca pre výpočet priznaného príspevku uznaným športom na Slovensku.

4.2.1 Vyššia veková kategória

Športovci sa môžu pohybovať medzi jednotlivými kategóriami nielen po dovŕšení veku definovaného na prechod do vyššej kategórie, ale aj skôr. Tento fakt predstavuje problém, ktorý sa bližšie rozoberie v časti *Prestup do iného zväzu*. Prechod do vyššej vekovej kategórie predstavuje **prirodzenú súčasť vývoja mladého športovca** a preto tento druh odchodu z organizovaného futbalu v SFZ v súťažiach s názvom obsahujúcim „LMŽ“ a „LSŽ“ nepredstavuje problém.

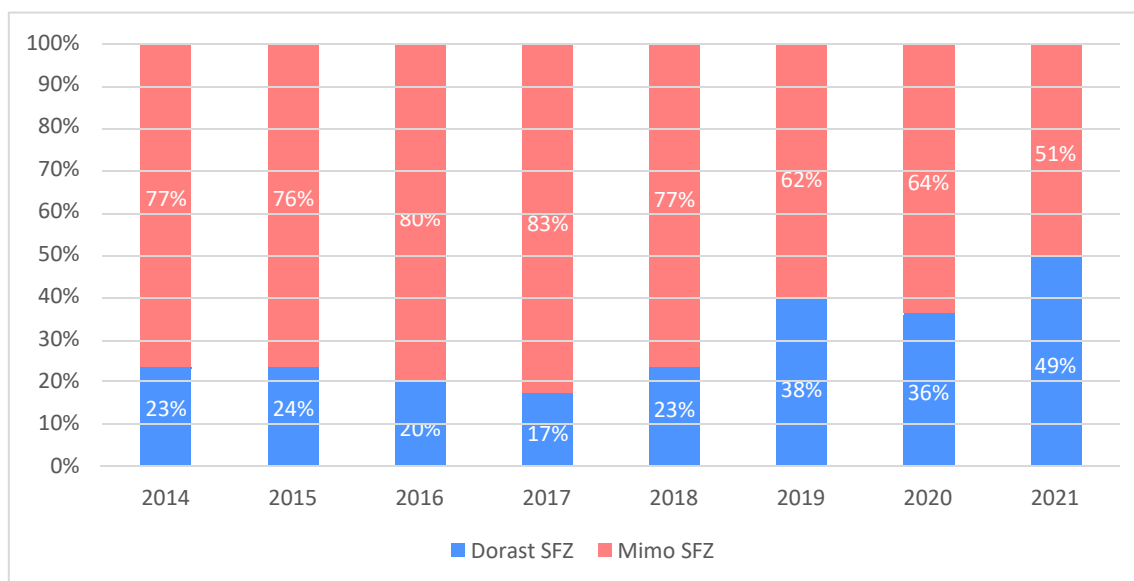
Tabuľka 4.2.2 ukazuje výsledky *Spracovanie dát*, kde sú zobrazené agregované štatistiky podľa sezóny, ktorá sa končila v roku, ktorý je uvedený z stĺpci rok.

Tabuľka 4.2.2 Výsledky pohybu kategórie žiaci v SFZ

Rok	Celkom	Pokračujú	Odišli	Dorast	Ž + D	Mimo SFZ
2014	2909	1793	1116	307	48	857
2015	3051	1927	1124	312	46	858
2016	2984	1854	1130	282	57	905
2017	3051	1810	1241	279	67	1029
2018	2894	1902	992	283	57	766
2019	3246	2139	1107	494	77	690
2020	3005	1923	1082	425	32	689
2021	2902	1989	913	620	176	469

Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Tabuľka 3.3.1 obsahuje stĺpce: **celkom**, ktorí predstavuje počet hráčov očistených od brankárov. V stĺpci **pokračujúci** sú hráči, ktorí ďalšiu sezónu nastupujú v súťažiach v SFZ, ktoré majú v názve „LMŽ“ a „LSŽ“. **Odišli** vyjadrujú výsledok odčítania pokračujúcich od celkom. **Dorast** predstavujú hráčov, ktorí ďalšiu sezónu nastupujú v súťažiach v SFZ s názvom „LDŽ“ a „LDŽ“. Stĺpec **Ž + D** hovorí o počte športovcov, ktorí ďalšiu sezónu nastúpili v oboch kategóriách (žiaci aj dorast). **Mimo SFZ** predstavuje počet športovcov, ktorí odišli odčítaných od počtu dorastencov prirátaný o počet Ž + D.



Graf 4.2.1 Kategorizovanie dôvodu odchodu športovcov – návrh
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Na *Graf 4.2.1* je zobrazený návrh sledovania dostupnej kategorizácie s dátami, ktoré poskytol SFZ. Návrh vychádza z *Graf 3.5.2*, ktorý tiež kategorizuje výstup – odchod z organizácie.

4.2.2 Prestup do iného zväzu

Prestupy v rámci zväzu, čiže medzi súťažami, ale aj mimo zväzu predstavujú faktor, ktorý poskytuje ďalšie údaje, informácie o odchode ako aj históriu hráča. Získaním štatistík o **migrácii športovcov** agregovane medzi zväzmi alebo aj detailne medzi súťažami sa **odhalenia trendy a nové, doteraz nedostupné možnosti** analyzovania, stratégií. Príkladom nových možností je napríklad sledovanie úspešnosti individuálnych politík zväzov, migrácia v jednotlivých súťažiach, alebo aj porovnanie migrácie obyvateľstva na Slovensku oproti migrácii športovcov vo futbale.

Hlavným problémom implementácie tejto dimenzie je však **spracovanie dát do potrebného stavu. Prestup**, resp. transfer sa vykonáva medzi organizáciami. Organizácia môže byť vo viacerých zväzoch a mať tímy vo viacerých súťažiach. K tomu je potrebné zohľadniť, že športovec nemusí prestúpiť do svojej vekovej kategórie, ale aj vyššej. To znamená, že na pohľad jednoduchý problém sa komplikuje a vzniká potreba komplexného riešenia. Po zvážení prínosov riešenia, ktoré prevyšujú náklady realizácie **sa riešenie vyžaduje**. Tento záver je na základe vyššie zmienených nových možností získania informácií, ktoré rozšíria vedomosti a znalosti v problematike odchodu z organizovaného futbalu, ale aj iných oblastiach medzi, ktoré patrí napríklad aj migrácia futbalistov.

Na *Obrázok 4.2.2* je zobrazená dimenzia futbal, ktorá vyjadruje zdroje prestupu do iného zväzu. Problematika prestupu sa dá riešiť primárne na dvoch úrovniach výstupu, a to **základné zväzy** alebo sledovanie prestupov do konkrétnych **oblastných zväzov**.

4.2.3 Zmena športu

Ďalšiu dimenziu odchodu z organizovaného futbalu zo *SFZ* predstavuje úplne opustenie futbalu, kde športovec prechádza na iný šport. Dôvodmi zmeny športu môžu byť faktory, akými sú napríklad aj **špecializácia** alebo **vyhorenie**, ktoré sú rozobrané v časti *Faktory vplyvajúce na odchod z organizovaného športu*. Zachytenie zmeny športu, resp. migrácie môže pomôcť pri spracovaní a následnom pochopení migrácie

medzi športmi a určiť trendy v mládežníckom športe, optimalizovaní stratégií alebo prerozdeľovaní finančných prostriedkov. Dáta, z ktorých sa vychádza pri dimenzií zmena športu, sú dostupné na **Slovenskom športovom portáli**.

Hlavný problém, ktorý nastáva pri spracovaní dimenzie zmeny športu, je **úspešné mapovanie hráčov**, ktorí sa nenachádzajú v predchádzajúcich dvoch dimenziách. Tento krok predstavuje nájdenie identifikátora na prepojenie dvoch individuálnych databáz. Identifikátor, ktorý sa používa v **Slovenskom športovom portáli** predstavuje kľúč **fyzickej osoby**, pričom **Slovenský futbalový zväz** využíva **číslo registrácie**.

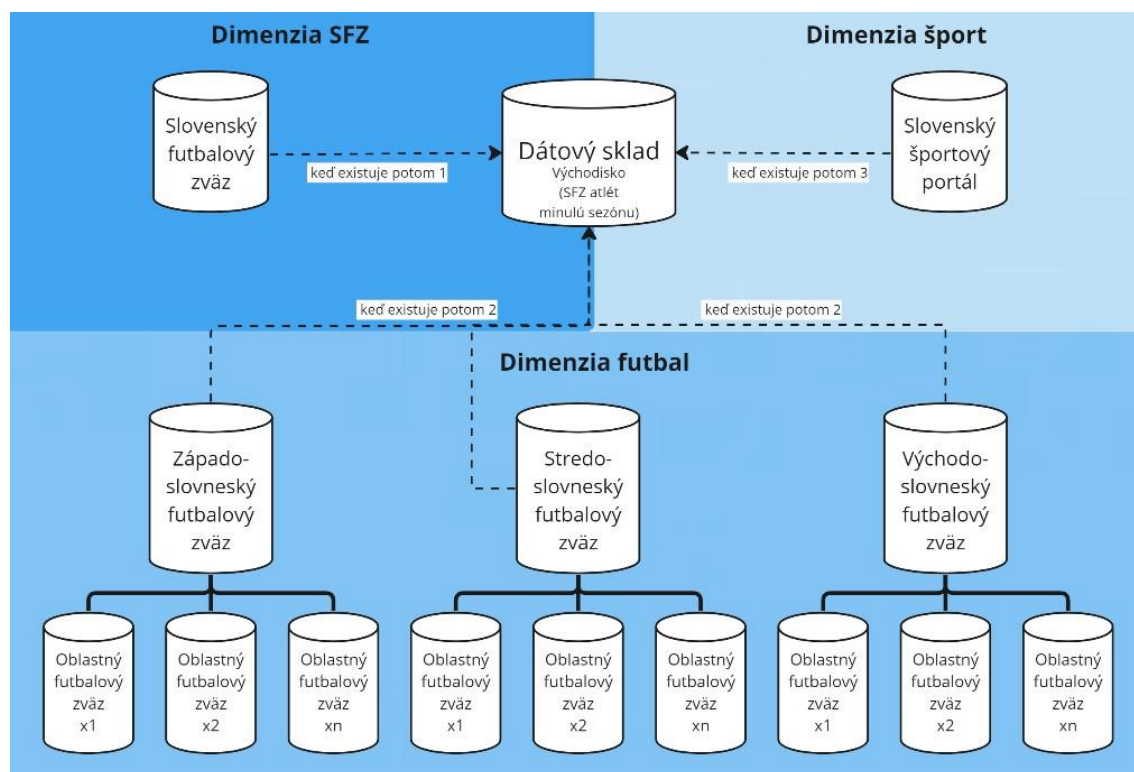
Na **Obrázok 4.2.2** je zobrazená dimenzia šport, ktorá predstavuje 3. krok počas hľadania športovca, kde východiskové podmienky sú, že sa nenachádzav predchádzajúcich dvoch dimenziách, SFZ a futbal.

4.2.4 Skončenie so športom

Všetci športovci, ktorí nie sú zaradení v žiadnej z uvedených dimenzií, sú kategorizovaní do skupiny **skončil so športom**, kde sa vychádza z predpokladu, že športovec nevykonáva žiaden iný šport. Zmena a skončenie so športom sa môžu vyjadriť len pre aktuálne sezóny z dôvodu nedostupnosti údajov pre športovcov, ktorí pokračovali určité čas s iným uznaným športom po skončení s organizovaným futbalom, ale aktuálne už nie sú evidovaní v žiadnom uznanom športe, a preto sa nenachádzajú v registri.

Skončenie so športom vyjadruje v návrhu tých, ktorí naozaj skončili, ale aj časť športovcov, ktorých sa nepodarilo úspešne zaradiť do predchádzajúcich dimenzií, kategórií. Nakoľko sa však jedná o oblasť športu, kde nie je dôležitá 100 % presnosť, prezentovaný návrh postačuje na vytváranie optimálnej stratégie riadenia udržateľnej športovej organizácie.

Na **Obrázok 1.2.1** je zobrazená ilustrácia návrhu dátových zdrojov, databáz pre realizáciu návrhu zvýšenia počtu dimenzií.



Obrázok 4.2.2 Návrh dátových zdrojov a postupu pre zvýšenie dimenzie výstupu
Zdroj: *vlastné spracovanie*

Východiskový bod **dátového skladu** predstavujú športovci, ktorí boli nominovaní v predchádzajúcej sezóne aspoň na jeden zápas v **kategórii žiak**. Takto vytvorený dataset s doplňujúcimi údajmi (štatistiky hráča, prípadne tímové výsledky za minulú sezónu) prechádza procesom kategorizácie výstupu – závislej premennej.

Na začiatku procesu kategorizovania sa kontroluje existencia športovca v **dimenzii SFZ** v **kategóriách žiak** alebo **dorast**. V prípade, ak sa športovec nenachádza v dimenzii SFZ, postupuje sa do **dimenzie futbal**, kde sa prechádzajú jednotlivé zväzy (do potrebnej, resp. zvolenej úrovne detailnosti). V prípade, ak sa športovec nachádza v jednom zo zväzov, priradí sa mu do **závislej premennej** (pokračuje) hodnota 2, prípadne pridá sa aj poznámka, v ktorom zväze pôsobí. Ak ani táto podmienka nie je splnená, pokračuje sa v procese kategorizácie a skúma sa športovcova existencia v Slovenskom športovom portáli. Ak sa podarí športovca spojiť s identifikátorom fyzickej osoby v Slovenskom športovom portáli, priradí sa hodnota 3, v prípade ak nie, ukončí sa **proces kategorizácie závislej premennej** a nastaví sa definovaná hodnota 0 (skončenie so športom).

4.2.5 Scenáre návrhu zvýšenia dimenzií výstupu

Cieľom časti zvýšenia dimenzií výstupu je **získať čo najväčšiu špecifikáciu závislej premennej**, odchodu z organizovaného futbalu v SFZ. To znamená klasifikovať športovcov a nájsť vlastnosti a atribúty skupiny. Zistením týchto informácií sa otvorí možnosť pre získanie nových vedomostí a znalostí, ktoré pozitívne ovplyvnia riadenie udržateľnej základne športovcov.

Tabuľka 4.2.3 Sumarizácia scenárov zvyšovania dimenzií výstupu

Variant scenáru	Dimenzia	Náročnosť	Očakávaný prínos	Pravdepodobnosť
1	SFZ	nízka	nízky až stredný	100 %
2	Futbal	stredná až vyššia	vysoký	75 %
3	Šport	vysoká	stredný	30 %

Zdroj: *vlastné spracovanie*

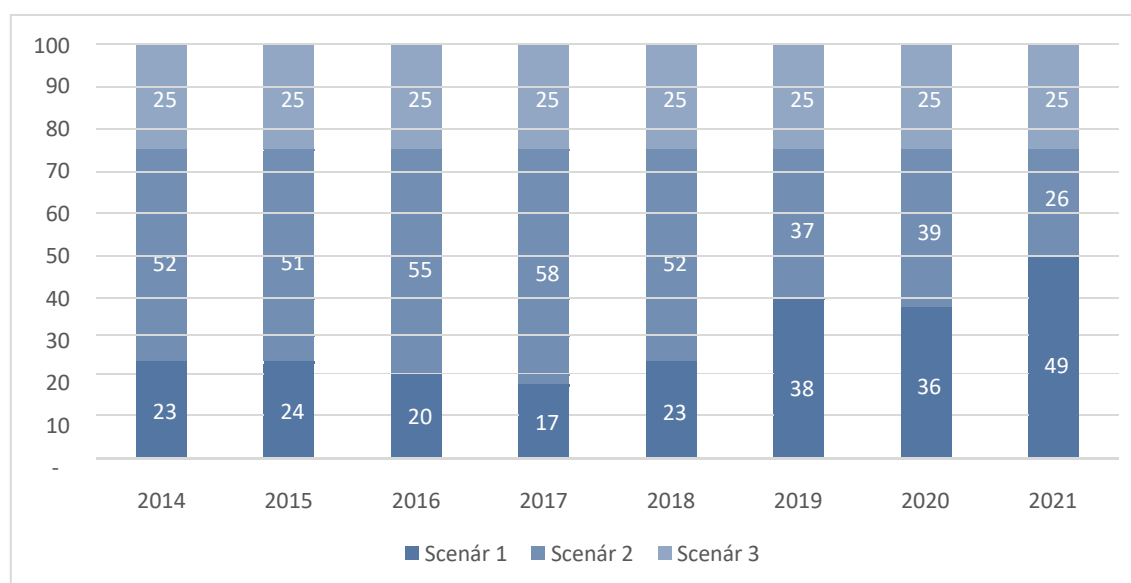
V *Tabuľka 4.2.3* sú zobrazené jednotlivé varianty pre návrh zvýšenia počtu dimenzií výstupu. Tieto varianty sú bližšie popísané v definícii konkrétnych dimenzií, kde sú uvedené aj jednotlivé problémy a limitácie. Pravdepodobnosť predstavuje odhad zrealizovania variantu scenáru, na základe analýzy. Každý ďalší variant obsahuje aplikáciu predchádzajúceho variantu. To znamená, že variant 2 predpokladá realizáciu variantu 1 a variant 3 realizáciu variantu 2.

5 PRÍNOSY A VYHODNOTENIE NÁVRHOV

Nasledujúca kapitola práce sa venuje ekonomickému a mimoekonomickému vyhodnoteniu návrhov diplomovej práce a jej prínosov. Špecificky sa zameriava na získavanie potrebných dát a zvýšenie dimenzie výstupu.

5.1 Mimoekonomické vyhodnotenie

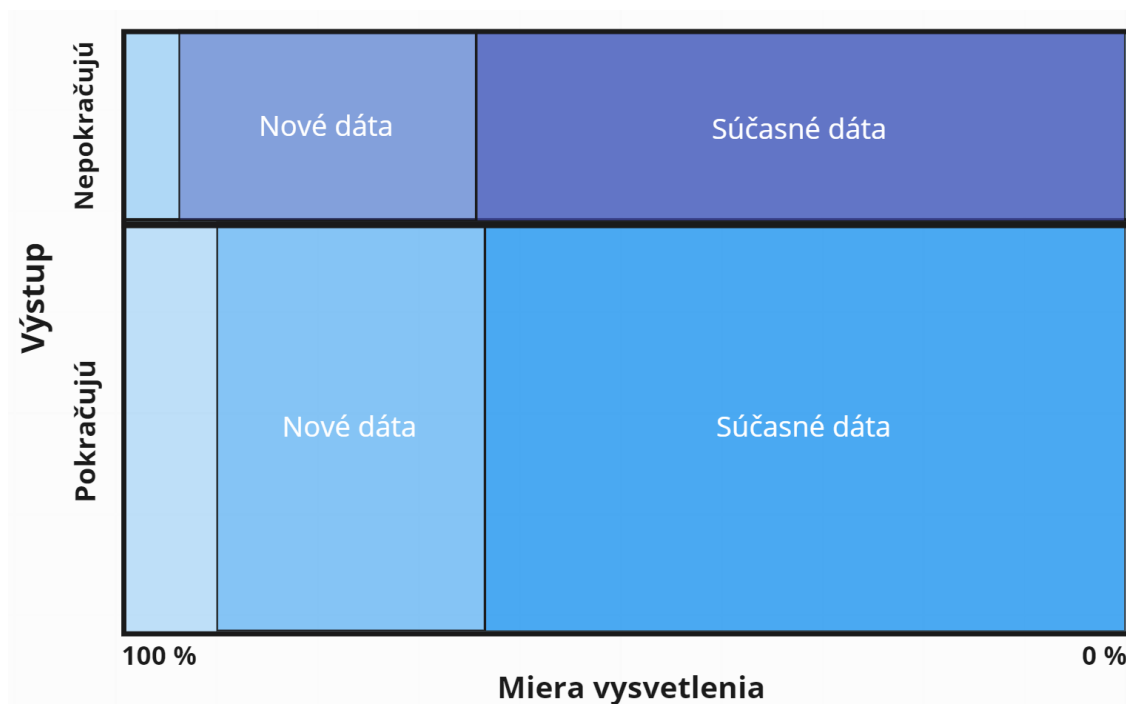
Mimoekonomické vyhodnotenie sa primárne zameriava na ukazovatele príčin odchodu. Správnym pochopením faktorov, ktoré najviac vplyvajú na skončeníu športovca s organizovaným futbalom, sa otvárajú možnosti efektívnejšej tvorby stratégie, tvorby rozpočtu, plánovania, organizácie a riadenia udržateľnej futbalovej základne, ktorá je dôležitá pre bezproblémový vývoj a evolúciu športu ako takého.



Graf 5.1.1 Predpokladaná miera kategorizácie odchodu podľa scenárov
Zdroj: *vlastné spracovanie na základe zálohy SFZ a API sportnet*

Na *Graf 5.1.1* je zobrazená miera vysvetlenia odchodu z organizovaného futbalu zo SFZ pri využití jednotlivých scenárov. Aplikovaním prezentovaného **návrhu zvýšenia dimenzie výstupu odchodu mládeže** sa zvýši miera a detailnosť pochopenia odchodu z organizovaného futbalu. Návrh otvorí nové možnosti analýzy a pohľadu na dáta, kde patrí napríklad vytvorenie **zhlukovej analýzy** podľa výstupu navrhovaného v scenároch, čo umožní zistenie nových informácií, vedomostí a znalostí v problematike futbalu na Slovensku.

S aktuálne dostupnými dátami so zálohou SFZ a sportnet API dosahuje najlepší klasifikačný model (SVM s využitím kernel triku, gamma 1 a cost 5) celkovú presnosť modelu na úrovni **0,648** (špecifická 0,658 a senzitivita 0,646). Realizáciou prieskumu formou dotazníku autor predpokladá vysvetlenie závislej premennej na úrovni 0,8 - 0,9 celkovej presnosti modelu, pri približne rovnakej miere vyváženosti modelu. Na *Obrázok 5.1.1* je graficky znázornený stav vysvetlenia problematiky aktuálne dostupnými dátami, v porovnaní s navrhovanými potrebnými dátami. Zobrazenie vychádza z Analytická časť práce.



Obrázok 5.1.1 Vysvetlenia dát – aktuálne vs navrhované dáta

Zdroj: *vlastné spracovanie*

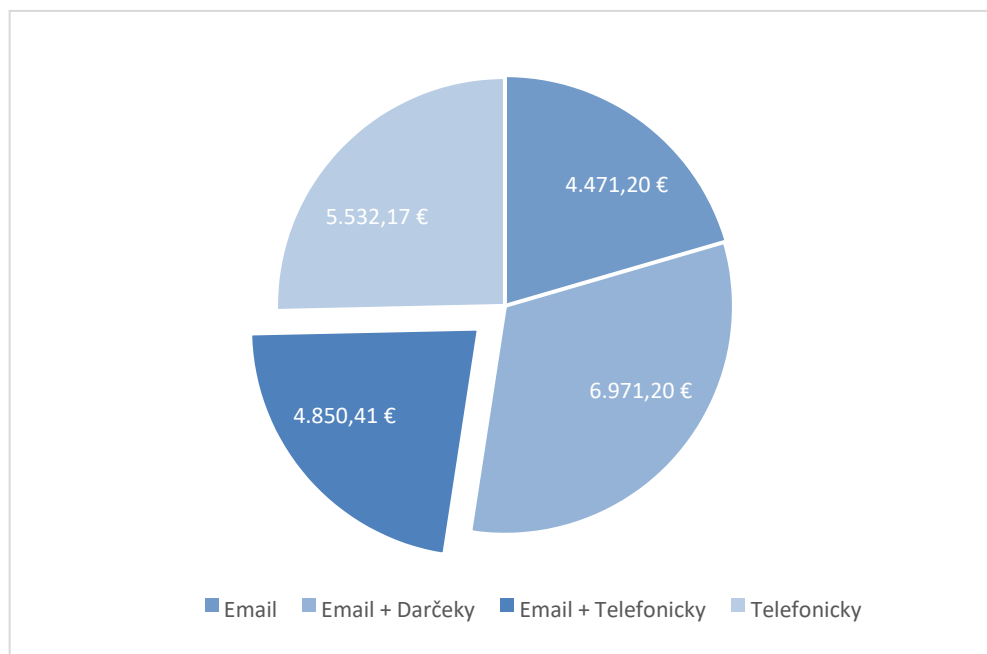
Pochopenie, prečo mládež odchádza zo športu, pomáha pri **kritických faktoroch riadenia športovej organizácie**. Najkritickejší faktor predstavuje **udržanie základne športovcov**. V prípade zvyšovania odchodu mladých športovcov môže dôjsť k spusteniu negatívnych vplyvov, ktoré ovplyvnia vývoj a budúcnosť futbalu. Porozumenie problematike otvára všetkým zainteresovaným subjektom (rodičom, trénerom, organizáciám, zväzom a pod.) možnosť k vytvoreniu vhodnej stratégie na **reguláciu odchodu**. Odhalenie faktorov, ktoré vplyvujú najviac na odchod môže viesť tiež k úpravám stratégie. Napríklad:

- **Zvýšenie bezpečnosti hráčov** – v prípade, ak najväčší faktor odchodu sú **zranenia**, organizácie sa môžu viac zamerať na zvýšenie bezpečnosti prostredníctvom zmien v pravidlách, zmeny tréningového procesu alebo vybavenia.
- **Podpora duševného zdravia** – v prípade, ak športovci opúšťajú organizovaný futbal z dôvodov, akými sú **tlak a vyhorenia**, zainteresované subjekty im môžu pomôcť nájsť balans, poskytnúť zdroje a podporu na vyrovnanie sa so športom a tak zostať v ňom.
- **Zvýšenie diverzity** – v prípade, ak mladí športovci obchádzajú z dôvodu **nedostatku rozmanitosti aktivít** počas vykonávania organizovaného futbalu, organizácie sa môžu zamerať na zmenu tréningového procesu alebo formátu súťaže.
- **Zlepšenie rozvoja zručností** – v prípade, ak sa športovci rozhodnú odísť z dôvodu **pocitu nedostatku príležitostí na rozvoj**, organizácie môžu vytvoriť programy prípadne iné príležitosti na rozvoj zručností a výkonu mladých športovcov, ktoré sú vytvorené na mieru.

Tieto príklady stratégií predstavujú len ukážku možností, ktoré prichádzajú so správnym pochopením vysvetľovacích faktorov a klasifikáciou odchodu mladých športovcov z organizovaného futbalu.

5.2 Ekonomické vyhodnotenie

V časti práce *ekonomické vyhodnotenie* sa odhadnú náklady na realizáciu návrhov, dopíšu sa najväčšie rizika a ohrozenia realizácie scenárov návrhov a taktiež sa definuje najväčší prínos práce. Odhady vychádzajú s dostupných informácií na internete, z vykonanej analýzy a z merania času potrebného na vypracovanie návrhov.



Graf 5.2.1 Náklady návrhu získania potrebných dát podľa scenárov

Zdroj: *vlastné spracovanie*

Na *Graf 5.2.1* sú zobrazené výsledné predpokladané sumy realizácie návrhu získania potrebných dát, kde výrez z koláčového grafu predstavuje odporúčaný variant realizácie. Bližšie informácie a údaje jednotlivých položiek nákladov sa nachádzajú v **prílohe B**.

Náklady sa skladajú z priemerných nákladov na hodinové stretnutie projektového manažmentu, ktoré sa pohybujú na úrovni 140 \$ za predpokladu, že sa stretnutí zúčastní 5 ľudí s priemerným platom 40000 \$ ročne. Ročný plat vychádza z priemernej mzdy a vyšších odvodových nákladov na Slovensku oproti USA. Všetky varianty majú rovnakú odhadovanú spotrebu času na projektovú prípravu. (hbr.org [online] 2016) Z fáz dotazníkového prieskumu: dizajnu, vytvorenia, testovania, vyplnenia, analýzy a reportovania. Pre vyčíslenie nákladov na jednotlivé fázy dotazníkového prieskumu sa používa odhadovaný strávený čas zamestnancov a náklady potrebných nástrojov na realizáciu. (platy.sk [online] 2023), (cliek4survey.sk [online] 2023) Konkrétne vyčíslenie jednotlivých položiek sa nachádza v **prílohe D**.

V *Tabuľka 5.2.1* sú zobrazené scenáre návrhu zvýšenia počtu dimenzií. Každý scenár ma **odhadnuté náklady**, ktoré vychádzajú z odhadu hodín potrebných na vypracovania v jednotlivých fázach realizácie návrhu. Nástroje sa do nákladov nepočítali, pretože podľa dostupných informácií, SFZ všetky technológie a softvéry potrebné pre realizáciu návrhu využíva. Optimistický a pesimistický odhad nákladov je vypočítaný z realistického odhadu, kde výsledok rozdiel je +/- 10 %. (platy.sk [online] 2023)

Tabuľka 5.2.1 Náklady návrhu zvýšenia počtu dimenzií podľa scenárov

Scenár - dimenzia	Optimistický	Realistický	Pesimistický
Scenár 1 - SFZ	2.449,49 €	2.721,66 €	2.993,83 €
Scenár 2 - Futbal	7.171,00 €	7.967,78 €	8.764,56 €
Scenár 3 - Šport	17.167,81 €	19.075,34 €	20.982,87 €

Zdroj: *vlastné spracovanie*

Bližšie informácie o vypočítaní nákladov návrhu zvýšenia počtu dimenzií výstupu sa nachádzajú v **prílohe C**.

Najväčším **rizikom** implementácie je nedodržanie rozpočtu pri realizovaní návrhu zvýšenia počtu dimenzií výstupu počas scenára 3. Riziko vzniká v dôsledku pracovania s **externými dátami**, kde nájdenie prepojenia medzi interným a externým ID športovcov môže zabrať veľa času a tak navýšiť predpokladané náklady. Ako nástroj na zníženie rizika môže pomôcť bližšia komunikácia so Slovenským športovým portálom. Na druhej strane návrh získania potrebných dát môže **ohroziť nesprávnosť kontaktných údajov**, respektíve **neaktuálnosť**. Toto ohrozenie môže nastať primárne pri prvom realizovaní dotazníkového prieskumu na celej vzorke od sezóny 2012/2013. V prípade vzniku definovaného ohrozenia dochádza k **navyšovaniu nákladov**, ktoré môžu byť až do výšky 100 % odhadovaných nákladov. Taktiež môže dôjsť k nedosiahnutiu stanovených cieľov dotazníkového prieskumu, z dôvodu **nesplnenia definovanej hodnoty alfy**.

Diplomová práca sa zaoberá problematikou odchodu mládeže z organizovaného športu. Uznané športy dostávajú finančné zdroje od rodičov športovcov vo forme poplatkov, od MINEDU ako príspevky, ale aj vo forme darov a dotácií od sponzorov. Pre zvyšovanie finančného kapitálu organizácií a dlhodobú udržateľnosť futbalu ako športu je potrebné aby základňa športovcov neklesala, ale naopak rástla. Preto práve pochopenie odchodu mládeže z organizovaného futbalu predstavuje kľúčovú oblasť pre **dosahovanie strategických cieľov** organizácií, zväzov a v neposlednom rade aj cieľov MINEDU. Práca problematiku úplne nevyriešila, ale posunula ju k riešeniu a poskytla odpovede na predtým nezodpovedané otázky. Príkladom týchto otázok môže byť: je schopne vysvetliť, respektíve predikovať odchod futbalistov na základe dát zo SFZ alebo ktoré faktory vplyvajú najviac na odchod z organizovaného futbalu. Na poslednú otázku bolo poskytnuté riešenie, ako odpoveď získať.

ZÁVER

Cieľom diplomovej práce bolo na základe teoretických východísk v oblasti športového manažmentu vypracovať návrh na zlepšenie súčasného stavu v špecifickej problematike, s implementáciou modelového riešenia. Ako špecifický problém bol vybraný odchod detí a mládeže z organizovaného futbalu v Slovenskom futbale po dovišení 15. roku života. Na problém poukázal a pomohol pri definícii vedúci IT oddelenia SFZ Ján Letko.

Na začiatku analýzy sa zisťovali faktory, ktoré podľa odbornej literatúry a článkov najviac vplyvajú na fenomén odchodu mládeže z organizovaného futbalu. Hlavným zistením bola veľkosť vplyvu, kde vnútorné prežívanie športovca, faktory zamerané na pocity a emócie, sú viac dôležitejšie ako objektívne hodnotenie. Taktiež mesiac a rok narodenia športovca pôsobí ako jeden z faktorov pri odchode zo SFZ. Analýza odbornej literatúry a článkov ukázala, že najlepšou metódou na získanie dát pre problematiku je dotazníkový prieskum. Pred vypracovaním modelového riešenia bolo nevyhnutné transformovať poskytnuté dáta od SFZ z objektivej formy do relačnej podoby. Po vykonaní dátovej transformácie sa doložili dáta, kde sa zistilo, že priemerný vek registrácie športovcov do SFZ každoročne klesá. Rozdiel v priemere registrácie je vidieť aj medzi pokračujúcimi a nepokračujúcimi športovcami, kde tí, ktorí pokračujú sa o 0,6 roku skôr registrujú do SFZ. Rovnako počet nominácií a odohratých zápasov zohráva kľúčovú úlohu vo vysvetlení odchodu z organizovaného futbalu. Faktor pozície sám o sebe nemá veľkú výpovednú hodnotu ale hráči, ktorí pôsobili na viacerých pozíciách majú vyššiu pravdepodobnosť pokračovania. Faktor individuálneho úspechu hráča, góly, pozitívne vplyva na zotrvanie v SFZ. Naopak udalosti, akými sú striedania, žlté karty a červené karty majú minimálnu výpovednú hodnotu v riešenej problematike. Analýza reportov šport v číslach priniesla zistenie, že počet trénerov pozitívne koreluje s počtom športovcov v CTM. Taktiež odhalila opačný trend migrácie športovcov na ZŠ zo západu na východ ako pri SŠŠ a ukázala zaujímavý spôsob klasifikácie futbalistov, ktorí odišli z CTM.

Výsledkom modelovania je zistenie, že aktuálne dáta SFZ sú nepostačujúce na vysvetlenie skúmanej problematiky. Z vytvorených a definovaných 13 premenných až 8 sa ukázalo štatisticky významných na hladine alfy 0,1, kde 6 premenných dosiahlo absolútnu hodnotu korelácie väčšiu ako 0,1. Najväčším problémom počas modelovania bola presnosť predikovania nepokračujúcich športovcov (špecificita), ktorá dosahovala vo vytvorených modeloch nízku úroveň. Z vytvorených klasifikačných modelov najlepšie výsledky dosahoval model SVM s využitím kernelu radial pri parametroch gamma 0,1 a cost 5. Výsledky hodnotenia modelu sú celková presnosť 64,80, špecificita 65,80 a senzitivita 64,60 percent.

Na základe vypracovanej analýzy sa vytvorili dva hlavné návrhy na zlepšenie aktuálnej situácie v odchode mladých športovcov z organizovaného futbalu v SFZ. Prvý návrh sa zameriava na spôsob získania potrebných dát na dostatočné vysvetlenie problematiky pomocou metódy dotazníkového prieskumu. Naopak druhý návrh sa zameriava na zvýšenie počtu dimenzií pohľadu na nepokračujúcich športovcov, kde sa pozerá aj na ďalšie pokračovanie športovcov po odchode z kategórie žiakov v SFZ. Konkrétne sa jedná o kategóriu dorast v SFZ, iné futbalové zväzy alebo iné športy.

Aplikovaním návrhov sa zvyšuje miera pochopenia odchodu mladých športovcov z organizovaného futbalu v SFZ. Realizáciou scenáru 2, návrhu zvýšenia počtu dimenzií odchodu sa zistí kam smerovalo 65 až 75 % športovcov po odchode z kategórie žiaci v SFZ, pri nákladoch 7 až 9 tisíc euro. Využitím návrhu získania potrebných dát, scenáru email + telefonicky, sa dosiahne zvýšenie miery vysvetlenia odchodu z aktuálnych 64,80 na celkovú presnosť modelu 80 až 90 % s podmienkou vyváženosti modelu. Okrem iného sa odhalia najväčšie faktory vplyvajúce na odchod zo SFZ, pričom realizácia návrhu vychádza v intervale od 4,5 do 5,5 tisíc euro.

Zoznam použitej literatúry

Tlačené zdroje

- [1] BIGELOW, B. a kol. 2016. *Youth Sports: Still Failing Our Kids – How to Really Fix It*. Bob Bigelow. 86 s. e-ISBN 978-1-3109-2118-6
- [2] FERNANDEZ, G. 2011. *Statistical Data Mining Using SAS Applications*. 2. vyd. Boca Raton: CRC Press, 477 s. e-ISBN 978-042-9131-87-5
- [3] FRIED, G. a kol. 2008. *Sport Finance*. 2. vyd. Human Kinetics. 4-18 s. e-ISBN 978-0-7360-6770-6
- [4] HYNDMAN, R. J. a kol. 2021. *Forecasting: Principles and Practice*. 3. vyd. Australia: Monash University. 422 s. e-ISBN-13 978-0987507136
- [5] KUHN, M. a kol. 2013. *Applied Predictive Modeling*. Springer. 443 s. e-ISBN 978-1-4614-6849-3
- [6] LAROSE, T. D. a kol. 2014. *Discovering Knowledge in Data: An Introduction to Data Mining*. 2. vyd. New York: John Wiley & Sons, 316 s. eBook ISBN 978-111-8874-05-9
- [7] PEDERSEN, P. M. a kol. 2017. *Contemporary Sport Management*. 6. vyd. Human Kinetics. 5-29 s. e-ISBN 978-1-4925-5096-9
- [8] SCHUMAKER, P. R. a kol. 2010. *Sport Data Mining*. Springer Science+Business Media, 477 s. e-ISBN 978-1-4416-6730-5
- [9] VEALEY, S. R. a kol. 2016. *Best practice for youth sport*. Human Kinetics. 117 s. e-ISBN 978-0-7360-6696-9

Elektronické online zdroje

- [10] AJILORE, O. H. a kol. 2021. *Interactive Survey Design Using Pidgin and GIFS* [online]. AfriCHI 2021: 3rd African Human-Computer Interaction Conference: Inclusiveness and Empowerment, 52-64 s. [citované 03.04.2023]. Dostupné na internete: <https://doi.org/10.1145/3448696.3448701>
- [11] AN YPEREN, N. W. a kol. 2022. *Predicting Dropout From Organized Football: A Prospective 4-Year Study Among Adolescent and Young Adult Football Players* [online]. FRONTIERS IN SPORTS AND ACTIVE LIVING [citované 01.04.2023]. Dostupné na internete: <https://www.webofscience.com/wos/woscc/full-record/WOS:000749619000001>
- [12] api.sportnet.online. 2023. *SportNet API docs* [online]. SportNet [citované 01.04.2023]. Dostupné na internete: <https://api.sportnet.online/v1/docs/#>
- [13] AZEVEDO, A. a kol. 2008. *KDD, semma and CRISP-DM: A parallel overview*. IN: Multi Conference on Computer Science and Information Systems [on-line]. 3/2008. p. 182-185. [citované 15.11.2022]. Dostupné na internete: https://www.scopus.com/record/display.uri?eid=2-s2.0-58449118356&origin=resultslist&sort=cp-f&src=s&st1=Crisp-dm&sid=ee359102f89a3d349461d7218f1e0e30&sot=b&sdt=b&sl=23&s=TITLE-ABS-KEY%28Crisp-dm%29&relpos=0&citeCnt=188&searchTerm=&featureToggles=FEATURE_NEW_DOC_DETAILS_EXPORT:1
- [14] BARQUERO-RUIZ, C. 2021. *A Teaching Games for Understanding Program to Deal With Reasons for Dropout in Under-11 Football* [online]. Research Quarterly for Exercise and Sport, číslo 92, problém 4, 618-629 s. [citované 16.11.2022]. Dostupné na internete: <https://doi.org/10.1080/02701367.2020.1759767>
- [15] BEI, Z. a kol. 2019. *Analysis of Factors Affecting Hit-and-Run and Non-Hit-and-Run in Vehicle-Bicycle Crashes: A Non-Parametric Approach Incorporating Data Imbalance Treatment* [online]. Sustainability, číslo 11. [citované 01.04.2023]. Dostupné na internete: <https://www.mdpi.com/2071-1050/11/5/1327>

- [16] click4survey.sk. 2023. *Cenová ponuka* [online]. Click4Survey s.r.o. [citované 05.04.2023]. Dostupné na internete: <https://www.click4survey.sk/ceny>
- [17] DEELEN, I. a kol. 2018. *Time-use and environmental determinants of dropout from organized youth football and tennis* [online]. BMC PUBLIC HEALTH, číslo 18 [citované 01.04.2023]. Dostupné na internete: <https://www.webofscience.com/wos/woscc/full-record/WOS:000441878700006>
- [18] DOHME, L. a kol. 2019. *Psychological Skills and Characteristics Facilitative of Youth Athletes' Development : A Systematic Review* [online]. Human Kinetics, 261-275 s. [citované 16.11.2022]. Dostupné na internete: <https://doi.org/10.1123/tsp.2018-0014>
- [19] DOS SANTOS, L. M. a kol. 2020. *Stress in Academic and Athletic Performance in Collegiate Athletes: A Narrative Review of Sources and Monitoring Strategies* [online]. Elite Sports and Performance Enhancement. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.3389/fspor.2020.00042>
- [20] education.msu.edu. 2019. *ISYS* [online]. Michigan state university. [citované 01.04.2023]. Dostupné na internete: <https://education.msu.edu/youth-sports/>
- [21] ERHART, M. a kol. 2009. *Effects of phone versus mail survey methods on the measurement of health-related quality of life and emotional and behavioural problems in adolescents* [online]. BMC Public Health, číslo 9. [citované 01.04.2023]. Dostupné na internete: <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-9-491>
- [22] GOMES, A.R. a kol. 2017. *Anxiety and burnout in young athletes: The mediating role of cognitive appraisal* [online]. SCANDINAVIAN JOURNAL OF MEDICINE & SCIENCE IN SPORTS, číslo 27, problém 12, 2116-2126 s. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.1111/sms.12841>
- [23] GOODFELLOW, I. a kol. 2016. *Deep Learning* [online] MIT Press. [citované 01.04.2023]. Dostupné na internete: <https://www.deeplearningbook.org/>
- [24] hbr.org. 2016. *Estimate the Cost of a Meeting with This Calculator* [online] Harward Business Review [citované 05.04.2023]. Dostupné na internete: <https://hbr.org/2016/01/estimate-the-cost-of-a-meeting-with-this-calculator>
- [25] CHAUDHURI, D. K. 2022. *Building Naive Bayes Classifier from Scratch to Perform Sentiment Analysis* [online]. Analytics Vidhya. [citované 01.04.2023]. Dostupné na internete: <https://www.analyticsvidhya.com/blog/2022/03/building-naive-bayes-classifier-from-scratch-to-perform-sentiment-analysis/>
- [26] CHEN, M.S. a kol. *Data Mining: An Overview from a Database Perspective* [online]. IN: IEEE Transactions on knowledge and data engineering, roč. 8, 2015, č. 6 [citované 05.11.2022]. Dostupné na internete: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=553155>
- [27] MARTINEZ-ALVARADO, J.R. a kol. 2021. *Burnout, Positivity and Passion in Young Mexican Athletes: The Mediating Effect of Social Support* [online]. INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH, číslo 18, problém 4. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.3390/ijerph18041757>
- [28] MENG-JIA, W. a kol. 2022. *Response rates of online surveys in published research: A meta-analysis* [online]. Computers in Human Behavior Reports, číslo 7. [citované 01.04.2023]. Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S2451958822000409>
- [29] MERKEL, D. 2013. *Youth sport: positive and negative impact on young athletes* [online]. Open Access J Sports Med. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.2147/OAJSM.S33556>

- [30] minedu.sk. 2022. *Alokácia finančných prostriedkov* [on-line]. MŠVVaŠ SR [citované 16.11.2022]. Dostupné na internete: <https://www.minedu.sk/alokacia-financnych-prostriedkov/>
- [31] MOESCH, K. a kol. 2012. *Late specialization: the key to success in centimeters, grams, or seconds (cgs) sports* [online]. SCANDINAVIAN JOURNAL OF MEDICINE & SCIENCE IN SPORTS, číslo 21, problém 6, E282-E290 s. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.1111/j.1600-0838.2010.01280.x>
- [32] mongodb.com, *What is MongoDB Compass?* [online]. MongoDB. [citované 01.04.2023]. Dostupné na internete: <https://www.mongodb.com/docs/compass/current/>
- [33] NGUYEN, G. a kol. 2019. *Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey* [on-line]. IN: Artificial Intelligence Review, roč. 8, č. 52, s. 77-124. [citované 14.11.2022]. Dostupné na internete: https://link.springer.com/article/10.1007/s10462-018-09679-z?utm_source=getftr&utm_medium=getftr&utm_campaign=getftr_pilot
- [34] PADHY, N. a kol. 2012. *The Survey of Data Mining Applications And Feature Scope* [on-line]. IN: International Journal of Computer Science, Engineering and Information Technology, roč. 2, č. 3 [citované 14.11.2022] Dostupné na internete: <https://arxiv.org/ftp/arxiv/papers/1211/1211.5723.pdf>
- [35] PANG, B. a kol. 2008 *Opinion Mining and Sentiment Analysis* [on-line]. IN: Information Retrieval, roč. 2, č. 1-2, s. 1-135. [citované 14-11-2023]. Dostupné na internete: <http://www.cs.cornell.edu/home/lee/omsa/omsa-published.pdf>
- [36] platy.sk/marketingovy-specialista. 2023. *marketing reklama pr* [online]. Profesia [citované 05.04.2023]. Dostupné na internete: <https://www.platy.sk/platy/marketing-reklama-pr/marketingovy-specialista>
- [37] platy.sk/telefonicky-operator. 2023. *administrativa* [online]. Profesia [citované 05.04.2023]. Dostupné na internete: <https://www.platy.sk/platy/administrativa/telefonicky-operator>
- [38] platy.sk/databazovy-analytik. 2023. *informacne technologie* [online]. Profesia [citované 05.04.2023]. Dostupné na internete: <https://www.platy.sk/platy/informacne-technologie/databazovy-analytik>
- [39] platy.sk/dotnet-programator. 2023. *informacne technologie* [online]. Profesia [citované 05.04.2023]. Dostupné na internete: <https://www.platy.sk/platy/informacne-technologie/dotnet-programator>
- [40] platy.sk/it-tester. 2023. *informacne technologie* [online]. Profesia [citované 05.04.2023]. Dostupné na internete: <https://www.platy.sk/platy/informacne-technologie/it-tester>
- [41] REIDER, B. 2017. *Too Much? Too Soon?* [online]. The American Journal of Sports Medicine, číslo 45, problém 6. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.1177/0363546517705349>
- [42] SALEH, A. a kol. 2017. *Examining Factors Impacting Online Survey Response Rates in Educational Research: Perceptions of Graduate Students* [online]. JDM, číslo 13, problém 29. [citované 03.04.2023]. Dostupné na internete: <https://files.eric.ed.gov/fulltext/ED596616.pdf>
- [43] SHARMA, S. 2021. *SVM: What makes it superior to the Maximal-Margin and Support Vector Classifiers?* [online]. Analytics Vidhya. [citované 01.04.2023]. Dostupné na internete: <https://www.analyticsvidhya.com/blog/2021/05/support-vector-machines/>

- [44] SCHLESINGER, T. a kol. 2018. *What is influencing the dropout behaviour of youth players from organised football?* [online]. German Journal of Exercise and Sport Research, číslo 48, 176-191 s. [citované 16.11.2022]. Dostupné na internete: <https://doi.org/10.1007/s12662-018-0513-4>
- [45] SJOERD, P. F. T. N. 2014. *Differences in response rates between mail, e-mail, and telephone follow-up in hand surgery research* [online]. Hand (N Y), 504-510 s. [citované 03.04.2023]. Dostupné na internete: <https://doi.org/10.1007/s11552-014-9618-x>
- [46] slov-lex.sk. 2021. *Zákon č. 440/2015 o športe a o zmene a doplnení niektorých zákonov* [on-line]. SLOV-LEX. [citované 16.11.2022]. Dostupné na internete: <https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2015/440/>
- [47] SMITH, R.E. a kol. 2007. *Effects of a motivational climate intervention for coaches on young athletes' sport performance anxiety* [online]. JOURNAL OF SPORT & EXERCISE PSYCHOLOGY, číslo 29, problém 1, 39-59 s. [citované 01.04.2023]. Dostupné na internete: <https://doi.org/10.1123/jsep.29.1.39>
- [48] SOLOMON, J. 2020. *Survey: Low-income kids are 6 times more likely to quit sports due to costs* [online]. Aspen Institute. [citované 01.04.2023]. Dostupné na internete: <https://www.aspenprojectplay.org/news/low-income-kids-are-6-times-more-likely-to-quit-sports-due-to-costs>
- [49] sport.gov.sk, *Register fyzických osôb v športe* [online]. Slovenský športový portal. [citované 01.04.2023]. Dostupné na internete: <https://sport.iedu.sk/Contact/Person/List>
- [50] statista.com: *Volume of data/information worldwide from 2010 to 2025* [on-line]. STATISTA. [citované 14.11.2022]. Dostupné na internete: <https://www.statista.com/statistics/871513/worldwide-data-created/>
- [51] statbank.dk. 2022. *Sports facilities by region, facility type and time* [on-line]. STATISTICS DENMARK [citované 16.11.2022]. Dostupné na internete: <https://www.statbank.dk/3592>
- [52] STROBEL, T. a kol. 2021. *Digital transformation in sport: The disruptive potential of 2 digitalization for sport management research* [online]. Sport, Business & Management, pp. 1-10. [citované 01.04.2023]. Dostupné na internete: https://boris.unibe.ch/153886/1/SBM_SI_Digital_transformation_Editorial_Manus_kriptform.pdf
- [53] Šport v číslach. 2021. [online] cvtislr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtislr.sk/buxus/docs/OddMladezASport/Sport/SportVCislach/2021/2021_sport-v-cislach.pdf
- [54] Šport v číslach. 2020. [online] cvtislr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtislr.sk/buxus/docs/OddMladezASport/Sport/SportVCislach/2020/2020_sport-v-cislach.pdf
- [55] Šport v číslach. 2019. [online] cvtislr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtislr.sk/buxus/docs/OddMladezASport/Sport/SportVCislach/2019/2019_sport-v-cislach.pdf
- [56] Šport v číslach. 2018. [online] cvtislr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtislr.sk/buxus/docs/OddMladezASport/Sport/SportVCislach/2018/2018_sport-v-cislach.pdf

- [57] Šport v číslach. 2017. [online] cvtistr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtistr.sk/buxus/docs//OddMladezASport/Sport/SportVCislach/2017/2017_sport-v-cislach.pdf
- [58] Šport v číslach. 2016. [online] cvtistr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtistr.sk/buxus/docs//OddMladezASport/Sport/SportVCislach/2016/2016_sport-v-cislach.pdf
- [59] Šport v číslach. 2015. [online] cvtistr.sk. [citované 01.04.2023]. Dostupné na internete: https://www.cvtistr.sk/buxus/docs//OddMladezASport/Sport/SportVCislach/2015/2015_sport-v-cislach.pdf
- [60] thesoccersidelines.com. 2019. *Relative Age Effect in Youth Soccer* [online] The Soccer Sidelines. [citované 01.04.2023]. Dostupné na internete: <https://thesoccersidelines.com/relative-age-effect-in-youth-soccer/>
- [61] tibco.com. 2022. what is a random forest [online]. TIBCO. [citované 01.04.2023]. Dostupné na internete: <https://www.tibco.com/reference-center/what-is-a-random-forest>
- [62] WAGNSSON, S. a kol. 2021. *Lessons Learned from a Multi-Level Intervention Program to Reduce Swedish Female Floorballers' Dropout Rate* [online]. Journal of Sport Psychology in Action, číslo 12, problém 4, 226-224 s. [citované 16.11.2022]. Dostupné na internete: <https://doi.org/10.1080/21520704.2020.1850576>
- [63] Záloha databázy Slovenského futbalového zväzu od sezóny. 2012-2022.
- [64] ZHONGBO, B.i a kol. 2021. *Sports Big Data: Management, Analysis, Applications, and Challenges* [online]. Cognitive Computing Solutions for Complexity Problems in Computational Social Systems. [citované 14.11.2022]. Dostupné na internete: <https://doi.org/10.1155/2021/6676297>

Zoznam príloh

Príloha A Dotazník s potrebnými informáciami.....	77
Príloha B Náklady návrhu získania potrebných dát podľa scenárov.....	78
Príloha C Náklady návrhu zvýšenie počtu dimenzií podľa scenárov	79
Príloha D Platy jednotlivých pozícií a skupín	80

Prílohy

Príloha A | Dotazník s potrebnými informáciami

Príklad otázky	Druh odpovede	Poznámka
Email (alebo iné ID)	Vloženie textu	identifikácia športovca
Ako dlho ste hrali/hráte organizovaný futbal?	Výber z možností	interval od do a iné
V akom veku ste začali s org. futbalom?	Vloženie textu	
Najvyššie dosiahnuto vzdelanie?	Výber z možností	študujem a ďalšie
Aktuálne povolanie?	Výber z možností	odvetvia + možnosti iné
Koľko iných športov ste vykonávali okrem org. futbalu?	Výber z možností	len športy na org. úrovni
Ako veľký pocit radosti ste vnímali z org. futbalu?	Škála od 1 po 10	
Ako veľký pocit zábavy ste vnímali z org. futbalu?	Škála od 1 po 10	
Ako ste vnímali okolie v rámci org. futbalu?	Škála od 1 po 10	okolia(tréneri, rodičia,...)
Ako veľmi ste vnímali futbal ako povinnosť?	Škála od 1 po 10	
V mládežníckych kategóriách ste mali oproti rovesníkom fyzicky a psychicky?	Výber z možností	výhody - nevýhodu
V mládežníckych kategóriách ste boli oproti rovesníkom výkonnostne?	Výber z možností	lepší - horší
V mládežníckych kategóriách ste vnímali váš tím ako?	Výber z možností	úspešný - neúspešný
V mládežníckych kategóriách ste vnímali tréningový proces ako?	Výber z možností	od kvalitný po nekvalitný
V mládežníckych kategóriách ste vnímali trénerov ako?	Výber z možností	od kvalitný po nekvalitný
V mládežníckych kategóriách ste vnímali intenzitu tréningov ako?	Výber z možností	nedostatočné nadmerné
V mládežníckych kategóriách ste vnímali frekvenciu tréningov ako?	Výber z možností	nedostatočné nadmerné
V akom veku ste sa začali špecializovať na futbal?	Výber z možností	nezačal, interval od do
Vnímali ste tlak pri org. futbale zo strany okolia?	Výber z možností	nie až áno, približne 5
Aké boli vaše očakávania od org. futbalu počas mládežníckych kategórií?	Výber z možností	Jedná možnosť je „iné“, vloženie textu
Ku ktorej z týchto dvoch oblastí ste sa viac prikláňali?	Škála od 1 po 10	škola až šport
Koľko zranení z org. futbalu ste mali počas mládežníckych kategórií?	Výber z možností	Možnosti „žiadne“, „iné“ a intervaly od do
Ako ste vnímali balans medzi org. futbalom a inými aktivitami?	Škála od 1 po 5	nerovnováha rovnováha
Ako ste vnímali počet príležitostí v org. futbale?	Škála od 1 po 5	veľa druhý málo
Ako ste vnímali kvalitu zariadení a vybavenia v org. futbale?	Škála od 1 po 5	dostatočná nedostatočná
Čo pre vás predstavuje najväčší faktor odchodu z org. futbalu?	Výber z možností	„iné“, ďalšie možnosti definované faktory

Zdroj: *vlastné spracovanie na základe analytickej časti*

Príloha B | Náklady návrhu získania potrebných dát podľa scenárov

Názov	Merná jednotka	Cena MJ	Scenár 1		Scenár 2		Scenár 3		Scenár 4	
			Počet	Spolu	Počet	Spolu	Počet	Spolu	Počet	Spolu
Projektový manažment	PM/h	128,00 €	20	2.560,00 €	20	2.560,00 €	20	2.560,00 €	20	2.560,00 €
Dizajn	MŠ/h	11,83 €	16	189,28 €	16	189,28 €	16	189,28 €	8	94,64 €
Vytvorenie	MŠ/h	11,83 €	24	283,92 €	24	283,92 €	24	283,92 €	16	189,28 €
nástroj	mesiac	39,00 €	2	78,00 €	2	78,00 €	2	78,00 €	0	- €
Testovanie	MŠ/h	11,83 €	10	118,30 €	10	118,30 €	7	82,81 €	5	59,15 €
respondenti	ks	5,00 €	10	50,00 €	10	50,00 €	5	25,00 €	0	- €
Náklady vyplnenia	MŠ/h	11,83 €	40	473,20 €	40	473,20 €	30	354,90 €	20	236,60 €
darčeky	ks	5,00 €	0	- €	500	2.500,00 €	0	- €	0	- €
volanie	TO/h	9,30 €	0	- €	0	- €	60	558,00 €	180	1.674,00 €
Analýza a reporting	DA/h	14,37 €	50	718,50 €	50	718,50 €	50	718,50 €	50	718,50 €
Spolu				4.471,20 €		6.971,20 €		4.850,41 €		5.532,17 €

Zdroj: vlastné spracovanie na základe Prílohy C

Príloha C | Náklady návrhu zvýšenie počtu dimenzií podľa scenárov

Názov	Merná jednotka	Cena MJ	Scenár 1		Scenár 2		Scenár 3	
			Počet	Spolu	Počet	Spolu	Počet	Spolu
Projektový manažment	PM/h	128,00 €	10	1.280,00 €	40	5.120,00 €	100	12.800,00 €
Analýza	DA/h	14,37 €	8	114,96 €	10	143,70 €	50	718,50 €
Analýza	DB/h	16,90 €	8	135,20 €	20	338,00 €	30	507,00 €
Návrh	DB/h	16,90 €	4	67,60 €	8	135,20 €	40	676,00 €
Návrh	C#/h	18,59 €	4	74,36 €	8	148,72 €	20	371,80 €
Nasadenie	DB/h	16,90 €	24	405,60 €	32	540,80 €	50	845,00 €
Nasadenie	C#/h	18,59 €	8	148,72 €	16	297,44 €	32	594,88 €
Testovanie	QA/h	13,52 €	8	108,16 €	24	324,48 €	40	540,80 €
Implementácia	DB/h	16,90 €	4	67,60 €	24	405,60 €	32	540,80 €
Implementácia	C#/h	18,59 €	4	74,36 €	8	148,72 €	32	594,88 €
Analýza a reporting	DA/h	14,37 €	10	143,70 €	16	229,92 €	24	344,88 €
Analýza a reporting	DB/h	16,90 €	6	101,40 €	8	135,20 €	32	540,80 €
Spolu			98	2.721,66 €	214	7.967,78 €	482	19.075,34 €

Zdroj: *vlastné spracovanie na základe Prílohy C*

Príloha D | Platy jednotlivých pozícií a skupín

Kód	Názov pozície	Merná jednotka	Hrubá mzda	Cena práce	Náklady na hodinu
MŠ	Marketingový špecialista	mesiac	1.400,00 €	1.892,80 €	11,83 €
DA	Dátový analytik	mesiac	1.700,00 €	2.298,40 €	14,37 €
TO	Telefonický operátor	mesiac	1.100,00 €	1.487,20 €	9,30 €
QA	IT tester	mesiac	1.600,00 €	2.163,20 €	13,52 €
DB	Databázový analytik	mesiac	2.000,00 €	2.704,00 €	16,90 €
C#	.net programátor	mesiac	2.200,00 €	2.974,40 €	18,59 €
PM	Projektový tím	5 členov / h			128,00 €

Zdroj: *vlastné spracovanie na základe platy.sk [36],[37],[38],[39],[40] a (hbr.org [online] 2016), (click4survey.sk [online])*