



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
Katedra biomedicínské informatiky

Vytvoření automatického reportu v R pro vyhodnocení výsledků DEMAT testu

Creation of an automatic report in R for the evaluation of DEMAT test results

Bakalářská práce

Studijní program: Informatika a kybernetika ve zdravotnictví

Studijní obor: Biomedicínská informatika

Autor bakalářské práce: Jan Pecha

Vedoucí bakalářské práce: Christiane Malá, Dipl. Ing (FH)

Kladno 2024

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Pecha** Jméno: **Jan** Osobní číslo: **510182**
Fakulta: **Fakulta biomedicínského inženýrství**
Garantující katedra: **Katedra informačních a komunikačních technologií v lékařství**
Studijní program: **Informatika a kybernetika ve zdravotnictví**

II. ÚDAJE K BAKALÁŘSKÉ PRÁCI

Název bakalářské práce:

Vytvoření automatického reportu v R pro vyhodnocení výsledků DEMAT testu

Název bakalářské práce anglicky:

Creation of an automatic report in R for the evaluation of DEMAT test results

Pokyny pro vypracování:

Cílem bakalářské práce je implementovat systém, který automaticky vytváří reporty z výsledků testů DEMAT žáků základních škol. Bakalářská práce je ve spolupráci s Technickou univerzitou v Drážďanech, která vědecky dohlíží na pilotní projekt „UniSchule Dresden“. V práci se seznámte s formátem datových souborů, které jsou k dispozici, a popíšete požadavky vědců v Drážďanech ohledně obsahu a vzhledu reportu. V rámci teoretické části práce se zabývejte samotným projektem a prozkoumejte vliv na studijní úspěšnost studentů. Poté pomocí dostupných datových sad vytvořte program v programovacím jazyce R, který je schopen report vytvořit. Dodržujte požadavky vědců z Drážďan ohledně tisku referenčních údajů, statistických výpočtů, tabulek a grafických znázornění. Implementujte řešení pro propojení dvou datových sad, které obsahují na jedné straně výsledky testů DEMAT a na straně druhé parametry nesouvisející přímo s učením (dotazník o sebereflexe). Odfiltrujte zvláště dobré (DEMAT>60) a zvláště slabé studenty (DEMAT<40) a vytvořte přizpůsobený report pro tyto skupiny.

Seznam doporučené literatury:

- [1] NOVÁKOVÁ, Kateřina a Petr VESELÝ, Jazyk R a tvorba grafů, Grada, Knihovna programátora, 2021, ISBN 978-80-271-3137-2
- [2] HOTHORN, Torsten a Brian EVERITT, A handbook of statistical analyses using R. Third edition, ed. Third edition, CRC Press, 2014, ISBN 978-1-4822-0458-2
- [3] Langner, A., Ritter, M., & Pesch, M., Das Reallabor Universitätsschule Dresden - forschungsmethodische Grundlagen., 2020-08-14, <https://doi.org/10.4119/pflb-3613>

Jméno a příjmení vedoucí(ho) bakalářské práce:

Dipl. Ing. Christiane Malá

Jméno a příjmení konzultanta(ky) bakalářské práce:

Datum zadání bakalářské práce: **14.02.2024**

Platnost zadání bakalářské práce: **20.09.2025**

doc. Ing. Karel Hána Ph.D.
vedoucí katedry

prof. MUDr. Jozef Rosina, Ph.D., MBA
děkan

PROHLÁŠENÍ

Prohlašuji, že jsem bakalářskou práci s názvem „Vytvoření automatického reportu v R pro vyhodnocení výsledků DEMAT testu“ vypracoval samostatně a použil k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k bakalářské práci. Prohlašuji, že tato práce je přesnou kopií odevzdané elektronické verze.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů.

V Kladně 15.05.2024

.....

Jan Pecha

PODĚKOVÁNÍ

Velmi rád bych poděkoval vedoucí této práce, inženýrce Christiane Malé, za výborné vedení veškerých projektů a této bakalářské práce. Její neustálé navrhování příležitostí mi velmi zpestřilo studium a načerpal jsem také mnoho zkušeností z jiných oborů.

V neposlední řadě děkuji své rodině za neustálou podporu při psaní této práce a při celém studiu. Chtěl bych věnovat samostatné poděkování mé sestře, která tuto práci pravopisně a sémanticky kontrolovala.

ABSTRAKT

Vytvoření automatického reportu v R pro vyhodnocení výsledků DEMAT testu

Hlavním cílem této práce je vytvoření programu, který dokáže upevnit strukturu prezentace výsledků testů DEMAT, které absolvovali žáci základních škol. Projekt vznikl v rámci mezinárodní spolupráce s Technickou univerzitou v Drážďanech, která na tyto testy dohlíží. V rámci této práce bylo nutné seznámit se s projektem, který probíhá na této škole a lépe mu porozumět z teoretického hlediska. Základ programu tvoří programovací jazyk R, jehož vývoj probíhal v prostředí RStudio. K práci s daty a zpřístupnění moderních přístupů k prezentaci a statistickému zpracování výsledků se použily různorodé dodatečné balíčky obsahující specifické funkce. Výsledný report obsahuje referenční údaje spojené s hodnoceným subjektem, tabulku se statistickými daty a grafická znázornění subjektu v akademickém kontextu a případně i jeho osobního hodnocení sebe sama a školy.

Klíčová slova

Programovací jazyk R, databáze, DEMAT automatizace tvorby reportu, Universitätsschule Dresden, analýza dat

ABSTRACT

Creation of an automatic report in R for the evaluation of DEMAT test results

The main goal of this work is to create a program that can consolidate the structure of the presentation of the results of DEMAT tests taken by primary school students. The project was developed as a part of an international collaboration with the Technical University of Dresden, which oversees the mentioned tests. As part of this work, it was necessary to become familiar with the project being carried out at this school and to understand it better from a theoretical point of view. The basis of the program is the R programming language, which was developed in the RStudio environment. Various additional packages, containing specific functions, were used to work with the data. This enabled the application of modern approaches for presentation and statistical processing of the results. The resulting report contains reference data associated with the subject being assessed, a table with statistical data and graphical representations of the subject in an academic context and, where appropriate, their personal evaluation of themselves and the school.

Keywords

R programming language, database, DEMAT automation of report creation, Universitätsschule Dresden, data analysis

Obsah

1	Úvod	4
2	Přehled současného stavu.....	6
2.1	Koncept Unischule Dresden	6
2.2	Sledování studentských výsledků na Unischule Dresden	9
2.3	Vliv okolních faktorů na výkon žáka ve škole	10
2.4	Problematika zpracování dat	16
2.5	Programovací jazyk R	18
3	Cíle práce	20
4	Technický a grafický návrh programu.....	21
4.1	Datasety	21
4.2	Rozdělení a koncept programu.....	23
4.3	Referenční údaje subjektu	26
4.4	Statistická prezentace výsledků subjektu	28
4.5	Profildíagram.....	30
4.6	Grafické znázornění vnějších ovlivňujících faktorů	32
4.7	Optimalizace programu	35
5	Implementace a výsledky	37
5.1	Uživatelské rozhraní.....	37
5.2	Vzhled vytvořeného reportu.....	39
6	Uživatelská dokumentace.....	41
7	Diskuse	43
8	Závěr	44
	Seznam použité literatury	45

1 Úvod

Vzdělání populace je základním stavebním kamenem funkční a spořádané společnosti. Každá země a zároveň i škola přistupuje ke vzdělání svých občanů z různých úhlů pohledu. Mohou se lišit jak svou razantností, přínosem nebo samotným obsahem. Ministerstva školství, které mají na starosti zmíněné téma, přicházejí každý rok s novými úpravami školních systémů jejich zemí. Tyto úpravy mají za úkol pozitivně ovlivnit úroveň vzdělání ve státě. Jedinečnost školního systému zaručuje specifický a na míru uzpůsobilý průběh vzdělání mladistvých, z čehož můžeme vyvodit, že na světě existuje nepřeberné množství dat, které produkuje kvantum výukových postupů a jejich testování. Tato obsáhlá data musí být nejprve zpracována a následně transformována do informace, kterou lze zprostředkovat a využít k usnesení rozhodnutí, zdali je současný systém výuky přínosný a efektivní. Z něj dále mohou posuzovat jiné aspekty a faktory jako například repetici nebo interaktivitu výuky.

Dalším důležitým faktorem ke zkoumání v rámci vzdělání studentů se v dnešní době ukázal psychický stav a duševní zdraví. Obzvláště v dnešní době je důležité podporovat a sledovat duševní zdraví člověka z důvodu jeho vlivu na výsledky studentů v průběhu náročných období dospívání. Tento nevyzpytatelný vnitřní faktor ovlivňuje chování a psychiku každého. A u dospívajících žáků to platí několikanásobně více než u dítěte v prepubertálním věku, jelikož pubertální změny a stavy způsobují, že chování jedince je mnohem více nekontrolovatelné a mnohdy i výbušné. Studie také ukázaly, že ovládnutí zkoušeného předmětu a motivace úsudku se pozitivně vážou k lepším výsledkům studentů vyšších ročníků, načež známky studentů nižších ročníků se poutají pouze ke zkušenostem s ovládnutím předmětu jako motivační bod k jeho celkovému pochopení. [1]

Příkladem školy, která se zabývá výše uvedenou problematikou, je univerzitní škola v Drážďanech, jejíž tým výzkumníků přichází s revolučním individuálním přístupem k žákovi. Systematický přístup využívaný na této škole si zakládá na odhadu v širokém a zároveň velmi individuálním spektru matematických schopností tamních studentů za pomoci unikátních testů DEMAT (Deutscher Mathematiktest). Pro zpracování velkého množství dat, pocházejících z výsledků testů, se využívají nejrůznější nástroje, které však

nemusí být navrženy ke zpracování a analýze výsledků z individuálních škol a je proto nutné vytvářet vlastní programy, které jsou toho schopné.

Hlavním cílem této práce je vytvořit takový program schopný automatického vytváření dynamických reportů o testování a jednotlivých subjektech na škole.

První část této práce se zabývá výukou na univerzitní škole v Drážďanech a možnostmi a postupy, které se momentálně užívají ke shromažďování výsledků testů. Dále je probrána problematika vlivu vnějších faktorů, například psychických problémů na akademický prospěch studenta na škole, čímž se primárně škola zabývá. Poté následuje část práce nastiňující technický návrh a zároveň implementace jednotlivých technologií a algoritmů k vytvoření programu. Poslední část je výhradně zaměřena na podobu uživatelského manuálu, popisu interakce uživatele s rozhraním, diskuzi, a nakonec zhodnocení celé práce.

2 Přehled současného stavu

V rámci této práce bylo nutné se zabírat problematikou několika témat ke správnému porozumění a k vedení debaty o požadavcích z partnerské školy. Dále bylo nezbytné vytvořit přehled o již zavedených postupech a jejich teoriích, které následně pozitivně ovlivnily začlenění vlastního řešení v podobě efektivních algoritmů. Došlo také k urychlení tvorby funkcí, které by byl program schopný zprostředkovat.

2.1 Koncept Unischule Dresden

Tým výzkumníků na Technické univerzitě (TU) v Drážďanech v Německu přišel se zcela unikátním konceptem výuky na základní škole. Samotný systém se jmenuje Unischule Dresden a využívá se ve školních zařízeních, která jsou součástí akademické obce města. Systém je zaměřen na vývoj žáka namísto třídy a pomáhá tak rozvíjet jeho osobnost a dovednosti, ve kterých vyniká. Žák tímto způsobem získává pocit identity, kdy si uvědomuje svoji jedinečnost ve společnosti, a následně přebírá zodpovědnost za svůj rozvoj. Námět projektu si zakládá na tradici rozvoji školy na základě vědeckých poznatků. Jako vedlejší cíl si klade univerzitní škola být modelovou institucí pro intenzivní přípravu budoucích učitelů. [2]

Tento přístup k výuce obnáší cenné pedagogické a analytické možnosti. Základem tohoto přístupu jsou unikátní testy DEMAT, které podstupují žáci základních škol zařazených v projektu. Jedná se o matematické testy prověřující tři tematické okruhy – algebra, geometrie a praktická matematika. Poslední zmíněný okruh se vyznačuje svou unikátností ve školním kurikulu. Vyskytují se v něm otázky a příklady, které žák může pragmaticky uplatnit v životě (viz Kapitola 2.2).

Výuka na školách zahrnutých v projektu TU probíhá z velké části formou dialogu, čímž dochází k lepšímu porozumění mezi studentem a vyučujícím. Jisté psychické pouto hraje důležitou roli ve věku dospívajícího studenta, jelikož ukotvuje upřímnost a pravdomluvnost mezi žákem a učitelem. Budováním takových vztahů se boří konzervativní mýty o školách, kde se neklade důraz na vzájemné pochopení mezi oběma stranami. Důsledkem tohoto jevu tak dochází ke vzniku nezdravého prostředí pro rozvoj. Takový dialog se stává základem a důležitou složkou lidského vývoje [2]. Osvojením komunikačních dovedností může student lépe zvládat různé nepříznivé životní situace. Tyto dovednosti jsou nesmírně důležité v různých sociálních komunitách. Proto

Unischule Dresden již v první a druhé třídě zavádí dětský parlament, který představuje participativní strukturu vyspělé společnosti. Příslušní delegáti se každý týden schází, aby formulovali a předali požadavky jejich vrstevníků a s ostatními členy parlamentu vykomunikovali náležitá řešení ke zlepšení každodenního života na škole. Samozřejmě tato setkání a forma demokracie je pod dohledem učitelů a výzkumníků, v případě že dojde k neshodě nebo k problémům v komunikaci.

Z předešlého odstavce vyplývá, že pojmy učení a rozvoj jsou významově propojené. Nejdůležitější faktor tohoto projektu však hraje studentovo odhodlání a motivace k zamyšlení se nad sebe samým, k budování a realizaci jeho zálib či snů. Tyto pudy jsou samozřejmě podporovány v prvních stádiích výuky, avšak konečná iniciativa se musí objevit u žáka [2].



Obrázek 2.1 – místnost pro žáky určena k vývoji design-thinking ve studentských skupinách [3]

Důležitou roli při výuce hraje také prostředí, ve kterém se student i učitel nachází. Člověk si podvědomě spojuje příjemný a uspořádaný prostor s pohodou, relaxací a v konečném důsledku s motivací a inspirací. Žák i učitel se do příjemného prostředí rádi vrací čímž lze do určité míry snížit množství absence studenta. Takový prostor se vyznačuje tím, že nabízí prostředí jak pro soukromí, tak i publicitu, aktivitu a kreativitu, dále jak pro individualitu, tak i týmovou práci.

V těchto prostorech mohou vyučující zapojit všechny žáky v projektové práci nebo je rozdělit a individuálně se jim věnovat. Příklad takového prostoru určeného k rozvoji design-thinkingu lze zahlédnout na Obr. 2.1. V kontrastu s moderní a uzpůsobenou místností, ve které se koná výuka, se standardní místnost na základní škole (Obr. 2.2) jeví jako pravý opak.



Obrázek 2.2 - názorná ukázka místnosti, ve které se koná výuka v České republice [4]

2.2 Sledování studentských výsledků na Unischule Dresden

Studenti podstupují každý rok matematické testy DEMAT. Jedná se o standardizované testy, vytvořené za pomoci dat shromažďovaných po všech federálních státech Německa. Existuje několik formátů tohoto typu testu záviselí na ročníku, ve kterém se student nachází. Liší se komplexností a zároveň se zvyšuje úroveň některých praktických otázek, které testují studentovu vyspělost k jejich objektivnímu zodpovězení. Škola tyto testy pořádá k vyhodnocení schopností celé třídy a dále i jednotlivých studentů.

DEMAT test se skládá ze tří tematických okruhů, které mají za úkol prověřit studenta v jednotlivých okruzích a zjistit tak jeho silné a slabé stránky. Tímto způsobem lze například tvořit skupiny podobných studentů k jejich efektivnějšímu vzdělávání. Nebo naopak se můžou poskládat skupiny studentů s rozdílnými slabými a silnými stránkami a lze tak dosáhnout lepší komunikace ve třídě, jejímž hlavním cílem je vytvořit základ nebo zdokonalení komunikačních a vůdcovských dovedností.

Prvním okruhem je obecná algebra. V té se prověřuje matematická kompetence žáka ve formě již zavedených a etablovaných otázek, jejichž obtížnost se pohybuje v rozsahu od jednoduchých po velmi obtížné. Další okruh tvoří geometrie, která se podobá svojí strukturou a obtížností části algebraické.

Finálním a nejvýznamnějším okruhem testu je praktická matematika. Tato část obsahuje matematické otázky vycházející z reálného světa a problémů, které se řeší v každodenním životě. Odpovědi zaznamenané v této sekci testu jsou stejně jako ostatní podrobeny důkladné analýze, jejíž výsledky mohou přinést pedagogům a výzkumníkům potřebné informace o kvalitě výuky a k zavedení a úpravě profilů individuálních studentů.

Profily studentů jsou stěžejními nástroji ke sledování žáka v dlouhodobém kontextu jeho pobytu na škole. Lze sledovat jeho vývoj v oblasti akademických výsledků a také v samotných okruzích, které mohou lépe naznačit specializaci nebo budoucí práci, o níž by se mohl student ucházet. Dále se jedná o důležitou pomůcku pro pedagogy k případné změně jejich přístupu k dotyčnému žákovi.

2.3 Vliv okolních faktorů na výkon žáka ve škole

Vliv vnějších a vnitřních faktorů na výsledky studenta v jejich akademickém životě byl podroben již mnoha studiím a experimentům. Tyto faktory se mohou projevovat jako diskutabilní a je obtížné je vědecky spojit se studentovo výkonem, avšak většina studií podporuje hypotézu obhajující přímo úměrnou závislost mezi psychoekonomickými podmínkami studenta a jeho výsledky. Výsledky daných studií se jeví jako zásadní, jelikož pomáhají nalézt souvislost mezi duševním zdravím v dětství, dosaženém vzdělání a socioekonomickou pozicí v dospělosti jedince [5].

Každý jedinec vyrůstá v jiném prostředí, jež výrazně ovlivňuje rodina, chování a různé sociální návyky jedince. Tyto faktory lze pozorovat ve všech stádiích života člověka, ale valná většina moderních studií se zaměřuje hlavně na studenty základních, středních a vysokých škol, protože právě v tomto období se člověk nejvíce mění jak fyzicky, tak mentálně. Ze studie konané pod vedením Suutela et al. [6] lze vyvodit pozitivní korelaci mezi úrovní vzdělání v oblasti, kde žijí děti vyznačující se pubertálním chováním, a jejich průměrnou známkou ve škole. Studie také vykazuje kladný účinek psychického dospívání u dívek na jejich průměrnou známku. Tento účinek však není u chlapců zpozorován nebo je téměř nepatrný.

Další studie z roku 2019 od Gräf et al. [5] se zabývá vlivem problémů v oblasti mentálního zdraví na akademické výsledky žáků prvního ročníku základní školy. Základní zdrojová populace výzkumu vycházela z 2003 dětí, jejichž průměrný věk činil sedm let. Poruchy nebo psychické potíže dětí byly identifikovány za pomoci specifického dotazníku, který se aplikoval před nástupem dítěte do školy. Byl vytvořen na základě požadavků studie a dodatečně na základě výsledků ze zdravotní prohlídky předškoláka, pokud byly dostupné. V rámci studie bylo nutné doložit lékařské výpisy budoucích studentů k případnému doložení faktů a správnému usnesení závěru, že dítě opravdu trpí nebo trpělo mentálním postižením.

Regression model	Health problem	N ^a	Effect estimate β^b (95% confidence interval)	P value
M1: mental health problems, overall	Mental health problems (with or without physical chronic health conditions)	601	-0.98 (-1.35, -0.61)	<0.001
	Physical chronic health conditions only	274	0.13 (-0.30, 0.57)	0.543
	None	384	Reference	
M2: internalising and externalising mental health problems	Internalising mental health problems (yes vs. no)	359	-0.12 (-0.50, 0.27)	0.550
	Externalising mental health problems (yes vs. no)	391	-1.44 (-1.83, -1.05)	<0.001
M3: conduct problems and hyperactivity inattention problems	Conduct problems (yes vs. no)	126	0.76 (0.21, 1.31)	0.007
	Hyperactivity inattention problems (yes vs. no)	326	-1.96 (-2.36, -1.56)	<0.001

Obrázek 2.3 - tabulka obsahující odhady efektu jednotlivých zdravotních problémů na akademické výsledky studenta [3]

Studie ukázala, že děti s psychickými problémy dosahovaly méně uspokojujivé výsledky než subjekty z referenční kohorty ve stejné věkové kategorii (Obr. 2.3). Tato studie zároveň poskytuje informaci o externalizujících problémech v chování, tj. problémy, které nastanou a ovlivní určité prostředí jako například porucha pozornosti a mají vysoký podíl na zhoršených výsledcích žáka. O internalizujících problémech lze také usoudit, že negativně ovlivňují studium. Avšak z teoretického hlediska nelze říci, zda musí vždy psychická onemocnění negativně ovlivnit žákovo počínání ve škole. Školák trpící sociální úzkostí může vynaložit vyšší úsilí při plnění domácích úkolů, aby předešel jemu nepříjemné zvýšené pozornosti ve školním prostředí. Studie tak podporuje myšlenku, aby se co nejrychleji identifikovaly děti s mentálními poruchami a předešlo se zbytečnému negativnímu vlivu na studentský a akademický život jedince.

Podle studie od Van Zwieten et al. [7] hraje v celém životě člověka důležitou roli středoškolské vzdělání. Výzkum se podrobněji zabíral spojitostí mezi časem, kdy jedinec dokončil střední školu, obdobím a trváním nízkého socioekonomického statusu v dětství a akademickými úspěchy během navštěvování střední školy. Nadprůměrné výsledky při studiu se pozitivně pojí se sociálním postavením ve společnosti a podmínkami pro život dospělého člověka. Studie přistupuje k proměnné socioekonomický status (SES) jako k dynamické expozici, kterou nelze vyjádřit pouhým průměrováním v malém časovém intervalu, jak je tomu právě v jiných studiích. Studie používá kohortní systém užívající volně dostupná data z longitudinální studie dětí v Austrálii (LSAC). SES se kalkuloval za pomoci příjmu, dosaženého vzdělání a současného profesního stavu rodiče/ů. Status dítěte mohl být na základě prahové hodnoty zařazen do jedné ze dvou skupin: nízký a vysoký. Expozice výzkumu se udávala na základě socioekonomického stavu rodiny ve čtyřech periodách (4-5 let, 6-7, 8-9 a 10-11), kde v každém období probíhalo vyhodnocení stavu. Tato kontinuální hodnota se využila k porovnání socioekonomického stavu rodiny oproti jiným rodinám ve stejné vlně testování. Akademické výsledky byly zaznamenány za pomoci skóre z okruhů témat, které mají prověřit matematickou a čtenářskou gramotnost studenta prvního ročníku střední školy.

Model	Explanation of Model	SES Variables in Model	Mean Difference (95% CI) in Reading z-Score	Model Likelihood-Ratio Test
Fully saturated	All trajectories of low SES from 4–5 to 10–11 years of age are important in explaining Grade 7 reading.	All periods of SES and all interactions		n/a
Null	SES has no effect on Grade 7 reading.	None	n/a	$P < .001$
4–5-year-old critical period	Low SES at 4–5 years has a critical effect on Grade 7 reading.	Low SES at 4–5 years old*	–0.50 (–0.56 to –0.43)	$P < .001$
6–7-year-old critical period	Low SES at 6–7 years has a critical effect on Grade 7 reading.	Low SES at 6–7 years old*	–0.50 (–0.56 to –0.44)	$P < .001$
8–9-year-old critical period	Low SES at 8–9 years has a critical effect on Grade 7 reading.	Low SES at 8–9 years old*	–0.45 (–0.51 to –0.39)	$P < .001$
10–11-year-old critical period	Low SES at 10–11 years has a critical effect on Grade 7 reading.	Low SES at 10–11 years old*	–0.49 (–0.55 to –0.43)	$P < .001$
Sensitive period/relaxed accumulation	Low SES from 4–5 to 10–11 years has cumulative impacts on Grade 7 reading, but the effect of each time point may differ.	Low SES at 4–5 years old* Low SES at 6–7 years old* Low SES at 8–9 years old* Low SES at 10–11 years old*	–0.20 (–0.30 to –0.11) –0.18 (–0.28 to –0.09) –0.02 (–0.12 to 0.09) –0.22 (–0.32 to –0.13)	$P = .05$
Linear strict accumulation	Regardless of timing, a longer duration of low SES from 4–5 to 10–11 years has cumulative impacts on Grade 7 reading (the effect of each time point is the same). This accumulation is assumed to be linear.	Number of periods of low SES (linear 0–4)	–0.16 (–0.17 to –0.14)	$P = .02$
Categorical strict accumulation	As per linear strict accumulation, but accumulation is not assumed to be linear (eg, there may be a threshold effect).	Number of periods of low SES (categorical 0–4)		$P = .02$
		0	Ref.	
		1	–0.26 (–0.37 to –0.15)	
		2	–0.30 (–0.41 to –0.20)	
		3	–0.46 (–0.56 to –0.35)	
		4	–0.63 (–0.70 to –0.56)	

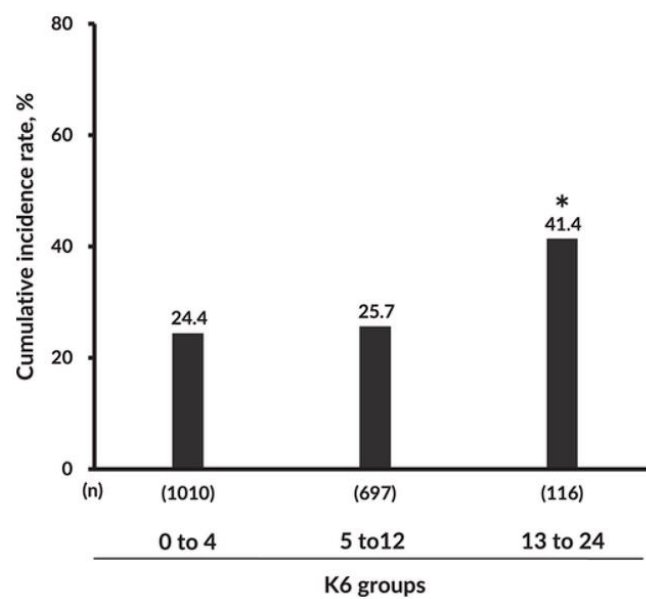
Obrázek 2.4 - tabulka o výsledcích studie zaměřené na efekt nízkého SES na z-skóre ve čtení v jednotlivých věkových skupinách [7]

Výsledky studie (Obr. 2.4) poukazují na vysokou kumulativní závislost mezi nízkým SES a zhoršenými výsledky u dětí ve věku od 4 do 5 let a 10 až 11 let. Obecně lze ze studie vyvodit záporný vliv nízkého SES rodiny dítěte na jeho výsledky a dosažené vzdělání, které se následně projevuje v dospělém životě s nepříznivými okolnostmi a blahobytem. [7]

Fenomén zhoršení výsledků v závislosti na mentálních problémech se objevuje nejen u studentů středních, ale i vyšších škol. Na světě se pozoruje zvýšené množství vysokoškoláků, kteří trpí zhoršeným duševním zdravím a tento problém se stává kritickým problémem veřejného zdraví. Naproti tomu Světová zdravotnická organizace (WHO) prováděla v roce 2016 celosvětové průzkumy, ve kterých bylo zjištěno, že pouze jedna pětina dotázaných studentů se pokusila hledat pomoc a adekvátní léčbu v oblasti mentálního zdraví. [8]

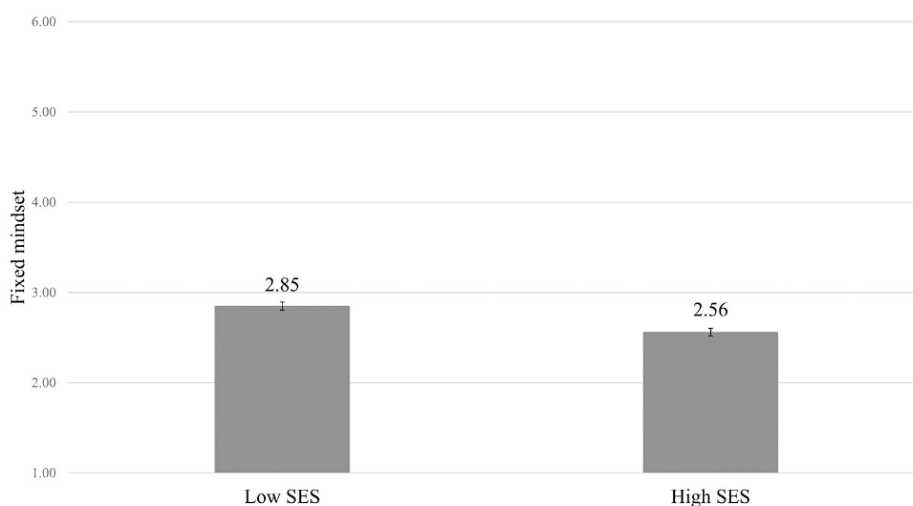
Studie od Chu et al. [8] se snažila zjistit závažnost tohoto problému u studentů bakalářských studijních programů v Japonsku. Sekundárním cílem této studie je určit influenci odlišných životních stylů na závislost mezi psychickým zdravím a špatnými akademickými výsledky. Přehled o výsledcích v tomto případě studie považovala jako skóre GPA, které se mění v průběhu studia v závislosti na získaných známkách. Posouzení psychického zdraví jedince bylo usneseno využitím Kesslerovy škály psychické újmy (K6), která obsahovala 6 položek k zodpovězení. Každá položka se mohla ohodnotit v rozsahu od 0 do 4 bodů, tedy celkové maximum bylo 24 bodů. Čím vyšší toto skóre nabývá hodnot, tím se zvyšuje podezření na přítomnost zhoršeného mentálního zdraví.

V rámci této studie se určily dva body k rozdělení studentů do tří kohort, každá zastupující jinou míru stavu psychického zdraví. K posouzení kvality životního stylu studenti samostatně odpovídali na průzkumové otázky přímo ve škole. Mezi rizikové faktory, použité k vyhodnocení chování mimo školu patřil čas strávený učením a hraním ve všední den, pitný režim, frekvence fyzické aktivity, návyk na každodenní přípravu snídaně, strava a čas probuzení se. Z Obr. 2.5 vyplývá narůstající riziko kumulativní incidence zhoršených výsledků u studentů, kteří ohodnotili otázky z dotazníku podle K6 vyššími hodnotami.



Obrázek 2.5 - graf rozdělení žáků do skupin podle dosažených bodů v K6 [6]

SES studentů hraje výraznou roli v různorodosti akademického života, tj. studentovo cesta k dospělosti, na jejímž konci lze vyvodit efektivitu akademického systému a jeho přínosy. Studie, která zkoumá téma vlivu socioekonomického postavení na studentovo způsob myšlení (*ang. Mindset*), probíhala pod vedením Destin et al. [9], který popisoval způsob myšlení jako čočku nebo brýle významně ovlivňující pohled člověka na svět. Mindsety mají tu schopnost změnit myšlenkový pochod a kompletně přetvořit identitu jedince. Tímto lze usoudit jejich nekompromisní podíl na výsledcích a vytrvání během studia. Jsou ovlivňovány mnoha faktory jako například finanční situace v rodině, městskou čtvrtí, mezilidskými kontakty a dalšími aspekty, které se vážou k socioekonomickému pozadí člověka. Protikladná úměra mezi druhem myšlení studenta a akademickým úspěchem je důkladně zdokumentována a popsána v článku od Blackwell, Trzesniewski, & Dweck, 2007; Yeager & Dweck, 2012. Vědci se shodli na dvou základních typech myšlení: první se nazývá jako statické, při kterém se dotyčnému v hlavě vytvoří mentální blok způsobující nemožnost rozvoje inteligence člověka. Naopak druhý typ, vývojový, charakterizuje inteligenci jedince jako schopnost, kterou je možné rozvíjet a neexistuje pro ni žádná horní hranice. Hlavní otázka, která byla studentovi kladena za účelem vyhodnocení úrovně socioekonomického stavu se týkala nejvyššího dosaženého vzdělání matky. Vhodnost otázky se vyvodila z jejího kontextu, který je konzistentně propojen s akademickými výsledky a účastníci mohou ve valné většině případů podat přesnou odpověď.



Obrázek 2.6 - vztah mezi úrovní SES a statickým typem myšlení [9]

Závěr této studie poukazuje na skutečnost, že studenti s vyšším SES mají nižší šanci si přisvojit statický způsob myšlení, a naopak ti s nižším statusem se k němu velmi často uchylují (Obr. 2.6). Také z finálního usnesení vzešlo výrazné propojení mindsetu s akademickým úspěchem ve formě přímé úměry. [9]

2.4 Problematika zpracování dat

Tento neobyčejný přístup k výuce na škole upřednostňující individuální rozvoj schopností žáka a zároveň vytváření mezilidských vztahů, však přichází se svými nevýhodami. Mezi ně patří nadměrná produkce nezpracovaných dat, která bez jakéhokoli dalšího zásahu nejsou zaměstnancům školy přínosná. Ke správnému předání a prezentaci výsledků je nutné nejdříve data předzpracovat. Tímto způsobem lze předejít chybám, jako například nesprávný formát, který není možné nijak interpretovat v kontextu ostatních výsledků.

Vznikající datasey, které se na úkor specifity výsledků mnohonásobně zvětšují ve svém objemu, se mohou do určité míry řadit vzhledem ke své strukturní komplexnosti do skupiny velkých dat (*ang. Big data*). Zmíněný typ dat se v přítomnosti využívá k podmanění moderních výzkumů, které se zabývají vším, co se děje na internetu v oblasti informatiky. Generovaná data mohou pocházet z pouhých transakcí provedených skrze online bankovníctví, emaily, počet kliků na videa až po zápisy ve zdravotnictví nebo mobilní telefony a jejich aplikace. K zúžitkování těchto dat je nutné, aby se provedl evoluční krok od tradičních metod datové analýzy. Základním postupem při pokusech pochopit tyto nadměrně objemná data je změna pohledu na ně, a to z hlediska tří hlavních komponent tvořící jejich jádro a strukturu: objem, variace a rychlost toku, resp. vzorkování dat. [10]

Datové soubory jakéhokoli formátu mohou obsahovat mnoho nesrovnalostí. Ve většině případů očekává program, který provádí analýzu, dokonalá data bez šumu a chybějících hodnot. Těmto problémům se však v reálném světě nelze vyvarovat a je nutné k nim aktivně přistupovat. V dnešní době existuje mnoho způsobů, jak si poradit se zmíněnými potížemi.

K redukci šumu se tvoří filtry s různými parametry, které se vzhledem k povaze dat mohou staticky i dynamicky měnit. Zavedením filtrů se lze vyvarovat nutnosti přeprogramovat celý proces získávání dat, což provádí jako možnost redukce šumu v datech. Přenastavení sběru dat se v tomto případě považuje změna klasifikace

jednotlivých záznamů ve formě zpřísnění vstupních kritérií, podle kterých se určují třídy dat. [11]

Chybějící hodnota označuje místo v jednom či více záznamech datasetu, kde nejsou obsažena žádná data. Existuje mnoho důvodů, proč se tato místa vytvářejí, přičemž většinou lze přičíst tuto anomálii selhávající funkčnosti zařízení sloužícího k získávání nebo ukládání dat. V opačném případě, kdy proces získávání dat není zcela zautomatizován, může vstoupit do hry lidský faktor, jenž se stává obtížně ovladatelným. K vyřešení problému chybějících dat se jeví využití datové regrese nebo parciální regrese nejmenších čtverců jako efektivní způsob odhadu dat pomocí modelů naučených na dostupných datech. [12]

V neposlední řadě mohou hodnoty v odlišných záznamech nabývat různých hodnot. Určité záznamy vychylující se z normy dat tak mohou drasticky zkreslovat výsledek analýzy, díky čemuž se celá studie znehodnotí. Tento fenomén lze prakticky vyřadit za pomoci normalizace dat. V datasetech se často vyskytují hodnoty užívající různé jednotky a intervaly, což může výrazně ovlivnit výkonnost a přesnost regresních modelů nebo jiných nástrojů užívaných ke statistickému vyhodnocení dat. Normalizace představuje klíčový aspekt v předzpracování dat, neboť umožňuje standardizovat hodnoty na jednotné měřítko, obvykle v rozmezí 0 až 1. Tímto způsobem se eliminují možná zkreslení vyplývající z diverzity rozsahů a procesy jsou tak schopny efektivněji generalizovat a pracovat nad různorodými daty. Normalizace představuje klíčový nástroj pro zlepšení výsledků statistických analýz velkých dat, které tak přispívají k vyšší úrovni porozumění.

Dalším důležitým úkonem při preparaci dat k analýze je redukce dat. Její metody mohou být rozděleny do 2 skupin podle zavedení různých algoritmů k šetření dat. První metoda redukuje početnou dimenzi dat, pokud k takové situaci dojde. U velkých dat se musí nejprve určit, jaké proměnné jsou při analýze bezprostředně nutné a nezaměnitelné. Lze tak drasticky redukovat složitost dat a předejít nárustu časové složitosti přes určitou hranici. Druhá skupina se zabývá diskretizací dat. Diskretizací lze za pomoci rozhodovacích stromů kategoricky rozřadit data, která jsou navíc přehlednější a jejich intervaly se nijak nepřekrývají. Rozřazením dat do intervalů vznikají ve stromě diskretizační body, díky kterým lze lépe číst, chápat a používat data uchovaná v diskretních hodnotách. [13]

2.5 Programovací jazyk R

První verze programovacího jazyku R byla vydána v roce 1993. Tento jazyk se ve světě programování a vývoje prosadil svojí schopností vykonávat náročné výpočetní úlohy, které by mohly trvat mnohem delší dobu, pokud bychom použili jiný programovací jazyk. Jeho poslední oficiální verze byla vydána v roce 2019. Přesto se používá díky komunitě plně nadšenců vytvářejících nové knihovny a balíčky funkcí, které rozšiřují schopnosti vývojového prostředí R.

Jeho kořeny zasahují až do programovacího jazyka S, přesněji řečeno do práce výzkumníka jménem John Chambers. Ten se svou firmou využil své prostředky na statistickou analýzu, jelikož v roce 1970 již byly stanoveny základy moderních technologií, jako například operační systém UNIX, nebo jazyky C a C++. Nestandardní přístup potřebný k různým typům dat si vyžadoval vytvoření nového a interaktivního prostředí k manipulaci s daty. Firma pod vedením Chamberse přinesla na trh jazyk S, neformálně „Statistics“. Následovaly roky úspěchu až k roku 1990, kdy byla vydána nová verze S-PLUS, která mohla být pořízena pouze jako komerční verze. Tímto krokem se sešli dva významní profesori zabývající se statistickou analýzou – Ross Ihaka a Robert Gentleman. Jejich střetnutí a seznámení dalo život a podobu jazyku R jak ho známe dnes. Rozhodli se vybudovat stabilní a intuitivní uživatelské prostředí, jehož absence u S byla velkou nevýhodou. Samotné jazyky nebyly odlišné z důvodu inspirace Chambersovo knihy o jazyku S, a ve vývojovém prostředí RStudio by bylo možné zkompileovat a spustit oba jazyky. Vydání R pod bezplatnou licenci se jevil jako úspěšný tah, jelikož v následujících letech se tento jazyk začal rapidně užívat k manipulaci a analýze velkých datových souborů. Po několika letech neustálého růstu popularity se zkušení programátoři a nadšenci do R rozhodli vytvářet takzvané knihovny obsahující nové dedikované funkce a rozšíření. V roce 1997 byl úspěšně vytvořen hlavní tým v čele s Ihaka, Chambers a Gentleman, na jehož sestavení se podílela celá komunita. Díky tomu jim byl udělen přístup k psaní zdrojového kódu R. Jako začátek nové éry statistického zpracování dat se založila a spustila oficiální stránka programovacího jazyka R. [14]

Všechny dodatkové komunitní knihovny se v zájmu udržení jednoduchosti a centralizace veškerého obsahu nacházejí v jednom ze tří největších online repositářích CRAN, který uchovává nejen balíčky, ale i samotný kompilátor a konzole ve formě programu RStudio. Základní technologické a výpočetní nástroje obsažené v základní instalaci však nejsou dostatečné pro požadavky dnešní moderní doby. Díky vlastním

knihovnám lze R využít od bioinformatiky přes různorodé zobrazovací techniky až po vývoj webových stránek. R se jeví jako ideální první programovací jazyk pro začátečníky díky jeho možnostem interakce mezi uživatelem a daty.

Tým pod R si zadal tři hlavní cíle, které má splňovat jejich produkt, z nichž první zajišťuje interakci s daty. Může jednat o předzpracování, normalizaci a celkové zajištění správné struktury dat před dalšími procesy. Druhým bodem je analýza zpracovaných k získání potřebných informací ze surových dat, jejichž užitečnost bez dostatečného opracování se chýlí k mizivým hodnotám. Poslední cíl se zabývá vizualizací získaných informací z dat. Vizualizačním technikám se v R nekladou meze, uživatel může zvolit a personalizovat veškeré druhy grafů či křivek, nebo použít jiné metody výstupů, například tabulku. [14]

3 Cíle práce

Cílem této práce je vytvořit program za pomoci programovacího jazyka R automatizující proces generování zpráv z výsledků speciálních testů DEMAT žáků základních škol. Tento projekt je prováděn ve spolupráci s Technickou univerzitou v Drážďanech, který vědecky a organizačně dohlíží na pilotní projekt „UniSchule Dresden“. Projekt se zaměřuje na několik klíčových oblastí, které jsou nezbytné k dosažení stanoveného cíle.

Prvním krokem je analýza dostupných datových souborů obsahující výsledky testů a jejich důkladná analýza. Po porozumění struktuře a formátu datasetů lze využít celý potenciál, který tato data mohou nabývat k vybudování programu generujícího potřebné zprávy. Přídavkem tohoto bodu je objasnění požadavků a vzhledu výsledného generovaného dokumentu vycházejícího z diskuze s vědci z Drážďan. Udržování stálé komunikace mezi oběma stranami zajišťuje, že výsledná zpráva bude odpovídat očekáváním a požadavkům na prezentaci výsledků testů DEMAT.

Další krok práce se zabývá teoretickým rámcem projektu univerzitní školy „UniSchule Dresden“. To zahrnuje detailní analýzu metodiky testů DEMAT a faktorů ovlivňujících studijní úspěch žáků. Porozumění teoretickému kontextu poskytuje značnou výhodu při analýze výsledků a tvorbě dokumentu v souladu s cíly a požadavky práce.

Hlavním cílem práce je implementace samotného programu za použití programovacího jazyka R, který dokáže automaticky generovat zprávy z příslušných datových souborů. Tento program bude schopen efektivně načítat a zpracovávat data, provádět statistické analýzy a generovat reporty na základě specifikací a požadavků výzkumníků z Drážďan. Součástí implementace je možnost propojení dvou datových sad a filtrování výsledků pro identifikaci akademicky zvláště dobrých a slabých žáků, což umožní generování přizpůsobených zpráv pro tyto skupiny.

Celkově bude práce přispívat k rozvoji projektu "UniSchule Dresden" a poskytne užitečný nástroj pro vyhodnocování výsledků testů DEMAT a jejich prezentaci ve formě zpráv.

4 Technický a grafický návrh programu

Proces tvorby softwaru byl rozdělen na několik tematických podkapitol, které podrobněji nastiňují práci s daty a funkcionalitu kódové části programu. Jejich seřazení odpovídá hierarchickému postavení v programu a ve výsledném reportu.

4.1 Datasetsy

Datové soubory, dále jen datasetsy, byly poskytnuty vědeckým týmem pod záštitou Univerzitní školy v Drážďanech. Souborů bylo hned několik, kdy každý uchovával výsledky jiných škol zařazených do projektu. Struktura a integrita jednotlivých souborů byla ucelená s drobnými výjimkami ve jménech sloupců, kdy v takových případech muselo dojít k jejich přejmenování za účelem správné funkčnosti softwaru.

Uspořádání dat v souborech kopírovalo následující strukturu: první sloupec vždy obsahoval anonymizovaný unikátní identifikátor studenta, k němuž se řadí daný záznam. Další sloupec uchovával data o příslušné ročníkové skupině (třídě), ke které byl na začátku ročníku přiřazen. Další sloupce obsahovaly informace o formě podstoupeného testu, bodová shrnutí, a nakonec bodové ohodnocení úloh v jednotlivých okruzích testu. (viz Obr. 4.1)

	A	B	C	D	E	F	G	H	I
1	identifier	Stammgruppe	Testform	FORM	Gesamtrohwert	PR_ges	PR_Band	T-Wert_ges	T-Wert_Band
2		13021 3a	DEMAT3+		13	19		41	
3		13022 3a	DEMAT3+		25	88		62	

A

	J	K	L	M	N	O	P
	Arithm_roh	Arithm_pr	Sachr_roh	Sachr_pr	Geom_roh	Geom_pr	ZS
	6		4		3		2
	11		8		6		3

B

	A	B	C	D	E
1	identifier	zufr_idx	swe_idx	motiv_idx	soz_herk_idx
2	7fee5	10	16	2	9
3	0abe3	14	12	4	10

C

Obrázek 4.1 - struktura listů v datasetu, A: výsledky testovacího období B: pokračování struktury datasetu z části A, C: odpovědi na otázky psycho- a socioekonomického počinání

Základním pilířem programu, který jakkoliv pracuje s daty, je jeho implementace algoritmu načítání samotných dat. Programovací jazyk R nabízí několik knihoven obsahujících dedikované funkce. V případě tohoto programu byl zvolen balíček *readxl*, jehož funkce se k načítání dat formátu Excel tabulky ukázala jako nejlepší volba ze všech možností. Při volání dané funkce bylo nutné specifikovat list, který měl program nalézt a načíst. (viz Kap. 5.1).

K předcházení neočekávaného chování programu se při načítání ošetřil případ prázdných buněk v tabulce. Při výskytu prázdné buňky se její hodnota přepíše jako NA (ang. Not available, čes. Nedostupný). S hodnotami NA už lze pracovat, jelikož jazyk R nabízí u valné většiny svých základních funkcí možnost nezahrnout tyto hodnoty, a díky tomu program nevytváří artefaktové hodnoty, které by mohly zkreslovat výsledky studentů na rozdíl od prázdných řetězců.

4.2 Rozdělení a koncept programu

K náležitému zachování přehledné struktury a k možné vizualizaci kódové části byla zvolena knihovna *rmarkdown* (verze 2.20), jejíž zavedení mění celý proces vytváření programu. Díky této knihovně dostává vývojář mimo jiné možnost rozdělit kód na neomezeného množství tzv. *chunků* (čes. region). Každý *chunk* se tak může zaměřovat na jinou funkčnost softwaru. Tato strategie tak zpřístupní dodatečné postupy při *debugování*, tedy odstraňování chyb a sledování průběžných výpočtů, a dokonce i generovaných grafů programu. V každém případě je stále nutné zachovat jistou hierarchii v pořadí jednotlivých částí, jinak se při tvorbě dokumentu, tj. spuštění celého kódu, můžou vyskytnout nečekané chyby a znemožnit generaci reportu. Jednotlivé *chunks* je možné libovolně pojmenovat, díky čemuž se uživatel lépe orientuje v kódu. Dalším důvodem, proč byla právě tato knihovna zvolena je možnost přenesení veškerých výsledků, u nichž lze nastavit, zda se ve finálním dokumentu zobrazí do jednoho ze tří formátů: Word, PDF nebo Html dokument. Každý se vyznačuje jinými výhodami a užitím v praxi. V kontextu této práce a k vyhovění požadavkům partnerské školy se dokumenty tvoří ve formátu PDF. Byl zvolen kvůli jeho zachování rozvržení obsahu a zároveň lze pomocí doprovodného softwaru měnit obsah bez větších potíží. Dále tento formát podporuje veškeré grafické komponenty jako například grafy či jiné obrázky. PDF je také možné otevřít na většině počítačů či mobilním telefonu.

Nultý *chunk* kódu zajišťuje ucelenou strukturu uživatelského rozhraní a možnou interakci uživatele s programem ještě před jeho spuštěním. Žádný kód se nespouští, a tak tento *chunk* nemá žádnou praktickou funkčnost. Jeho výstup nelze nikde v dokumentu zachytit, program pouze tyto proměnné načte do paměti a používá je jen v některých regionech kódu. V této části se upravuje vzhled dokumentu jako jsou okraje, formát nebo font. Po tomto nastavení následuje stanovení vstupních parametrů, které může uživatel volně měnit (viz. Kap. 5.1). Patří mezi ně například název školy, typ hodnoceného subjektu, id subjektu (pouze pokud se jedná o typ subjektu student), třída nebo ročník, forma testu a zvolení datasetu, ze kterého mají výpočty a dokument vycházet.

První region kódu se stará o speciální usazení parametrů v paměti, aby se zpřístupnily za běhu programu. V prvním *chunku* kódu vytvořeného v prostředí *rmarkdown* se jako je to u jiných programovacích jazyků, načítají, případně i stahují potřebné knihovny. V případě jazyka R lze využít knihovnu *pacman*, která je schopna právě výše zmíněný proces bez nutnosti opakovaného vypisování několika řádků kódu vykonat. Tato

knihovna však dokáže udržovat samotné balíčky aktualizované a při každém spuštění kontroluje dostupnost nových verzí načítaných knihoven.

Další chunk nazývajícím se „načítací“ má za úkol postarat se o prvotní zprostředkování dat. Jedná se o načtení datasetu do paměti a ošetření nečíselných a prázdných hodnot, které se mohou při vytváření Excel tabulky objevit, přičemž k tomuto jevu může dojít jak díky lidskému faktoru, tak chybě vyskytnuté při ukládání v použitém softwaru. V této části se také určuje typ hodnoceného subjektu, díky čemuž se uvolní místo v paměti.

Následující úsek kódu již produkuje viditelný výsledek v dokumentu. Jedná se o část, kde se zpracovávají referenční údaje o hodnoceném subjektu jako například identifikátor studenta, třída, ročník, název školy a forma testu. Zde se také poprvé uplatňuje knihovna *kableExtra*, jenž se využila kvůli její praktičnosti při manipulaci s daty k formátování a zkonstruování jednoduché struktury výsledné tabulky.

Další region se již stará o vytváření tabulky, která organizuje data v podobě výsledků testu z Excel tabulky. Nad daty se provádí statistické výpočty ve formě průměru:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \quad (\text{F4.1})$$

kde X_i je prvek záznamu, i zkoumaných dat a n je počet prvků v množině.

Nutnost jejich výpočtu však záleží na typu hodnoceného subjektu, který byl uživatelem zvolen. V případě, že se jedná o individuálního studenta, výsledky jeho testu jsou zkopírovány a bez dalšího zpracování přidány do datové struktury obsahující ostatní výsledky. Pokud je vybraným subjektem skupina nebo celá škola, nad daty se provádí u každého sloupce výsledků průměr, jehož hodnota se uloží do zmíněné struktury. Průměry se dále na požádání od vědců ze školy v Drážďanech zaokrouhlují na jedno desetinné nebo celé číslo. Nevýhodou načítaných dat se stala jejich struktura, protože samotné R tato data načítá bez téměř jakékoliv integrity ve struktuře, a proto bylo nutné vytvořit vlastní strukturu tabulky, do které se vyjmutá data vkládají. Je nutné zmínit skutečnost, kdy při různých formách testů se částečně mění názvy a struktura tabulky. Poslední úsek kódu, který není závislý na formě testu či typu subjektu, se stará o úspěšné vykreslení dané tabulky. Při vytváření je k dispozici mnoho možností jak esteticky upravit tabulku díky objemnému balíčku *knitr* obsahující mnoho funkcí a nástrojů ke customizaci vytvářených objektů v programu.

Následující chunk se týká dalšího požadavku a tím je graf znázorňující „Prozentranc“, jehož významu se věnuje kap. 4.5. Graf byl vytvořen pomocí funkce *ggplot* z knihovny funkcí *ggplot2*. Knihovna nabízí široké možnosti, jak přizpůsobit graf podle svých potřeb.

V posledním úseku kódu se provádí párování mezi dvěma listy ze stejného Excel sešitu pomocí unikátního identifikátoru studentů. Ke konstrukci grafu byl použit stejný přístup jako v předešlé části kódu. Avšak v této části proběhla důležitá úprava struktury dat k jejich zpracování. K tomu byla využita funkce *melt* z knihovny *reshape2*, která mění integritu v případě dlouhých řádků jednoho záznamu na krátký zápis pouze se třemi sloupci.

4.3 Referenční údaje subjektu

První tabulku vytváří třetí chunk kódu, jehož funkčnost byla podrobně vysvětlena v předešlé kapitole. Jednotlivé údaje jsou obsaženy ve vybraném datovém souboru. Jejich zapracování do dokumentu se může lišit v závislosti na zvoleném typu subjektu uživatelem v uživatelském rozhraní. Tyto údaje jsou prezentovány ve formě tabulky, která systematicky vyobrazuje dané informace (Obr. 4.2).

Personenbezogene Daten

Identifler	13038
Klasse	3a
Erhebungszeitraum	duben 2024
Vergleichswerte	—
Klassenstufe	—
Name der Schule	TU Dresden
Jahrgang	—
Testform	DEMAT3

Obrázek 4.2 - tabulka s referenčními údaji hodnoceného subjektu

První řádek tabulky (*Identifler*) poskytuje informaci o unikátním identifikačním řetězci, pomocí kterého lze spojit konečný dokument s dotyčným studentem školy. Tento řádek je přítomný pouze pokud je verze dokumentu nastavena na vyhodnocení studenta. V opačném případě se řádek vynechá z výpisu. Další řádek (*Klasse*) se využívá jako identifikátor třídy, kdy na daných školách jsou jednotlivým třídám přiřazeny unikátní názvy. Jedná se o praktické označení ke zrychlení procesu vyhledávání a organizaci života jak pedagogů, tak studentů v akademické obci. Řádek má i další důležitou funkci a tou je identifikace v případě, že uživatel programu zvolí jako typ hodnoceného subjektu třídu (skupinu). Lze tak jednoduše identifikovat, pro koho a z jakých dat byl daný report vytvořen. Následující řádek (*Vergleichswerte*) se může doslovně do češtiny přeložit jako „Srovnávací hodnoty“. Jedná se o experimentální hodnotu, která by se v praxi měla vyplňovat pověřenými pedagogy a podávat dodatečnou informaci o vytvořeném reportu ve formě několikáslovného ohodnocení generovaných výsledků. Další řádek (*Klassenstufe*) obsahuje informaci o úrovni třídy, kterou určuje systém a rada akademické obce. V neposlední řadě se v tabulce nachází jméno školy, které znovu mohou vyplnit pověření zaměstnanci škol při vytváření reportu. V rámci tohoto projektu operuje několik škol, které poskytly data k testování programu, a proto byla vytvořena tato možnost k umožnění úpravy a rozlišení dokumentu mezi školami. Předposlední řádek (*Jahrgang*) prezentuje možnost zapsání ročníku školy, avšak tato funkce nebyla v programu zatím

využita kvůli přítomnosti názvu třídy, po jehož vyhledání ve školní databázi se všechny potřebné informace dozví zaměstnanec. Posledním řádkem tabulky s referenčními údaji (*Testform*) odkazuje na formu testu, kterou podstoupil hodnocený subjekt. Forma testu je úzce spojena s ročníkem třídy a mezi jednotlivými verzemi dochází ke změně obtížnosti a povaze otázek.

4.4 Statistická prezentace výsledků subjektu

Čtvrtá část kódu je jedna ze základních pilířů programu, co se týče požadavků od TU, zodpovídá za vykreslení a formulování jak získaných, tak vypočtených dat do tabulky s hodnocením subjektu. Znovu se zde uplatňuje responzivní design, kdy program nejdříve zjistí, o jaký subjekt ze tří možných se jedná a na základě zvoleného typu se změní jak forma výpočtů hodnot, tak i struktura dat. Další rozhodnutí, které musí uživatel a pak i program učinit, je zvolení formy testu, kterou subjekt podstoupil (viz. Kapitola 5.1). Po tomto kroku začne program zpracovávat potřebná data z datasetu. Zpracovaná data se dále ukládají do pomocných proměnných ke zlepšení viditelnosti a čitelnosti kódu.

V této části také probíhají veškeré pomocné výpočty a statistická vyhodnocení. Po dokončení výpočtů se následně vykreslí tabulka zobrazující veškerá bodová ohodnocení u jednotlivých okruhů a oblastí přítomných v testu. Obr. 4.3 představuje vizuální formu druhé tabulky dokumentu, která obsahuje bodové počínání studenta a veškeré další ohodnocení vnímány jako nezbytné pro pedagogy.

Testergebnisse

Aufgabentyp		Rohwerte			
ZS	Zahlenstrahl	1			1
AD	Addition	2			2
SB	Subtraktion	1			1
MP	Multiplikation	1			1
SA	Sachaufgaben		3		3
SP	Spiegelung			3	3
FL	Formen legen			3	3
LS	Längen schätzen			1	1
LU	Längen umrechnen		4		4
	Subtest	Arithmetik	Sachaufgaben	Geometrie	Gesamtwert
	Rohwert (Summe)	5	7	7	19
	Prozentrang	71	70	32	51
	T-Wert				50

Obrázek 4.3 - tabulka s podrobnými výsledky testu [autor]

Nutno podotknout, že tabulky pro dvě další možnosti hodnocených subjektů, třídu a celou školu, jsou téměř identické. Vizuální podoba tabulky byla uzpůsobena požadavkům pracovníků z TU, kdy se tabulka rozdělila na 3 části. První část ve formě prvních dvou sloupců s názvem *Aufgabentyp* má za úkol popsat o jaké typy úloh se jedná. K ní se dále váže druhá část *Rohwerte* tvořící zbytek sloupců tabulky. Tato část se věnuje podrobnému výpisu výsledků k jednotlivým úlohám, jejichž povahu nastiňuje druhý sloupec poukazující na název úlohy. Je rozdělena na tři tematické sloupce, které rozdělují bodová ohodnocení do jednotlivých okruhů testu. Poslední sloupec slouží k revizi celkového bodového dosažení.

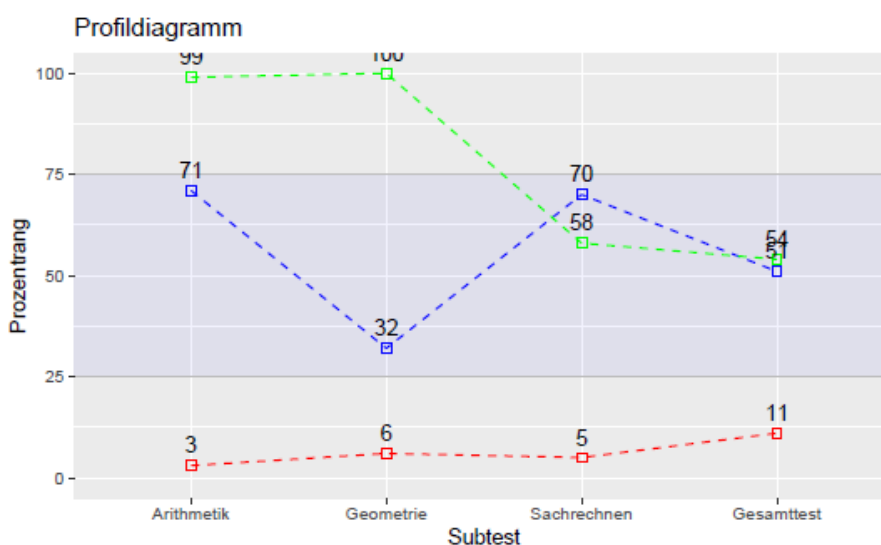
Třetí a poslední úsek tabulky podává informace o celkovém hodnocení, kupříkladu první řádek *Rohwert*, suma bodů v jednotlivých okruzích a bodů za všechny okruhy. Další řádek, *Prozentrang*, představuje studentovo usazení do kontextu celé třídy. Můžeme ho přirovnat ke specifitějšímu druhu percentilu (viz. Kapitola 4.5). Hodnota *T-Wert* se odvíjí od průměrné představy bodového ohodnocení žáka. Pokud se žák s touto hodnotou vychýlí z průměru je mu věnována zvýšená pozornost a také podrobení náležitým otázkám, které mohou odpovědět výzkumníkům na otázku proč se tak stalo z hlediska psychologického. Rozsah této hodnoty je určen pro každý okruh a pak i pro celkové ohodnocení. Získává se zprůměrováním výsledků všech žáků a z předešlého dlouhodobého testování společně s vědeckými podklady. Nachází se již vypočítaný v příložené Excel tabulce společně s datasetem.

4.5 Profildiagram

Následující komponenta dokumentu se prezentuje v podobě grafu a nese název *Profildiagramm*. Tato figura nastiňuje percentilové pořadí daného subjektu závislé na frekvenční distribuci, která je menší než jeho skóre. Vyjadřuje tedy relativní postavení vůči celkové populaci.

Funkcionalitu a přínos percentilového pořadí lze uvést na jednoduchém příkladu. Máme například čísla v rozmezí od 0 do 100, pokud se z nich vybere číslo 60, tak 40 procent čísel je na tom lépe, tj. jsou vyšší a naopak 60 procent čísel jsou nižší. V praktickém příkladě se může uvažovat o testu a podmínkách přijetí na určitou školu. Škola má stanoveno, že pro úspěšné přijetí na školu musí žák získat percentil 85. To znamená, že žák dosáhl lepších výsledků než 85 procent uchazečů. Z tohoto úsudku také vyplývá, že na školu je přijato pouze 15 procent ze všech uchazečů.

Vizuální podoba grafu percentilového pořadí, který lze vidět v dokumentu, se značně upravila oproti verzi grafu sloužícímu k obyčejné grafické prezentaci jednotlivých bodů. Změny se zavedly ke splnění požadavků z TU vzhledem k povaze testování žáka. Na Obr. 4.4 je na první pohled možné identifikovat patřičné úpravy. Největší z nich je zjednodušené zobrazení percentilu modrými spojovanými body a zároveň rozdělení grafu na čtyři úseky podle okruhů obsažených v testu. Na horizontální ose jsou vyneseny úseky a na vertikální se nachází percentil. V grafu je také vyznačena oblast modrou barvou, která představuje optimální střední hodnotu percentilu. Výzkumníci tento rozsah určili na základě jejich uvážení a předešlých výzkumů.



Obrázek 4.4 - náhled grafu znázorňujícího percentil a interval spolehlivosti

Dalším významným ukazatelem v grafu je vykreslení spodní a horní hranice percentilu příslušnými červenými a zelenými ukazateli. Informace o hodnotě obou hodnot se nachází v každém záznamu u žáka. Hodnoty představují spodní a horní interval spolehlivosti, který se počítá jako průměr hodnot plus a minus variace hodnot. Interval přibližuje odhad rozmezí, do kterého může výsledek žáka spadnout s určitou jistotou, pokud by test měl podstoupit znovu. Spolehlivost ve statistice je odlišný způsob k popisu pravděpodobnosti.

Jednotlivé body jsou spojeny přímkami, čímž je možné z grafu vyčíst mnohem více informací, například jak si žák vedl v každém tematickém okruhu oproti ostatním a jak velká je výchylka jeho úspěšnosti jak celkově, tak v individuálních okruzích.

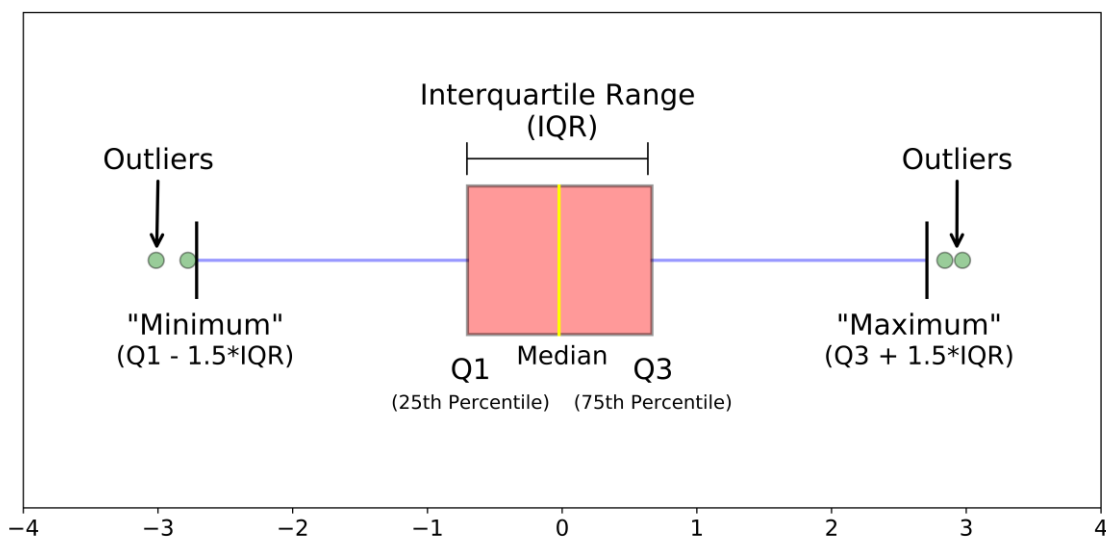
Ke grafu se pojí i přiložená legenda, která má za úkol popsat kohortní body a křivky rozdělené pouze barvou.

4.6 Grafické znázornění vnějších ovlivňujících faktorů

Poslední viditelnou část dokumentu sestavují výsledky ankety, která se týká psycho- a socioekonomických faktorů žáka. Každému studentovi je v určitých časových intervalech nabídnuta možnost vyplnit dotazník obsahující otázky na citlivé oblasti žáka. Odpovědi ve formě bodového ohodnocení se ukládají do stejného sešitu datasetu, ve kterém jsou uchovány i testové výsledky studenta. Bodové rozsahy, ze kterých mohou studenti volit, se mohou lišit v závislosti na okruhu otázek.

Pomocí jedinečného identifikátoru lze spojit studenta s výsledky a lze jeho odpovědi na anketu zobrazit v jeho reportu. Samotná prezentace výsledků nabývá speciální formy krabicového grafu (*ang. Boxplot*). V oblasti pro vykreslení se nachází celkem čtyři grafy, kdy každý z nich znázorňuje ohodnocení jednotlivých otázek. Při navrhování grafické stránky krabicového grafu se dbalo na kontrastní, a přesto příjemné barvy.

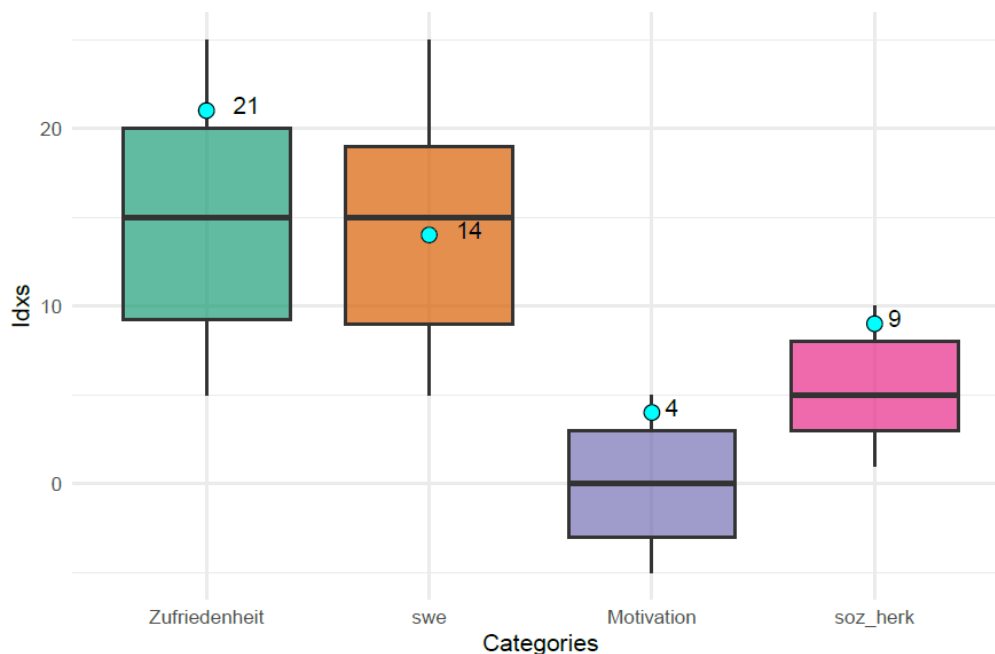
Krabicový graf (Obr. 4.5) je jednou ze standardních forem statistického a grafického znázornění a používá se k usnesení několika statistických hodnot o velkém množství dat. Skládá se z několika částí, které mohou poskytnout důležité informace o celku. Zobrazuje datový soubor a popisuje jeho rozsah vizualizací.



Obrázek 4.5 - ukázka krabicového grafu včetně popisu jeho částí (zleva – Odlehlá hodnota, minimum, percentil, rozmezí mezi kvartily, maximum) [15]

Inspekcí grafu odspodu lze zachytit první důležitý prvek poskytující informaci o minimální hodnotě souboru dat. Ukazatel minimální hodnoty se může s každým grafem lišit, kdy ve většině případů se používá vodorovná krátká příčka, také nazývaná „fous“. U grafu v dokumentu však nebylo nutné v zájmu přehlednosti volit zmíněnou možnost.

Údaj, který je možný vyvodit z grafu následující po minimální hodnotě, se vyobrazuje pomocí svislé nebo v některých scénářích i vodorovné úsečky. Tato úsečka určuje, jaké hodnoty se nacházejí v 25 procentech datového souboru. Říká se jí také kvartil. Kvartily rozdělují soubor dat do čtyř velikostních úseků, které dělí data podle jejich četnosti a výskytu v souboru. Na boxplotu (Obr. 4.6) lze v prvním sloupci znázorňujícím tematický okruh spokojenost vidět, že první kvartil dat je 9 a třetí kvartil 20.



Obrázek 4.6 - názorná ukázka boxplotu v reportu (Idxs – indexy, zelený – spokojenost (Zufriedenheit), oranžový - , fialový – motivace (Motivation), růžový - , tyrkysové body – body vybraného studenta)

Pohledem na širokou část grafu zjistíme, že medián okruhu se blíží k hodnotě 15, který se určuje podle rozdělující tučné čáry v této části. Tato část také dohromady spojuje druhý a třetí kvartil a tvoří tak 50 percentilovou část dat. Poslední a nejvrchnější částí grafu odhaluje maximum zpracovaných dat.

Verze boxplotu v programu byla dále obohacena o nový prvek na základě požadavků pracovníků TU. Jedná se o ukazatel přesné hodnoty, kterou daný žák zvolil v anketě. Mezi možné symboly pro znázornění bodového hodnocení vybraného žáka byl zařazen trojúhelník, hvězda a kolečko. Po konzultaci s odpovědnými pracovníky se kolečko zvolilo jako nejvhodnější možnost. V krabicovém grafu ho lze spatřit s tyrkysovou barvou, u nějž je vypsáno číslo reflektující odpověď v dané kategorii.

Dodatečně je nutno podotknout, že tato část dokumentu se vykreslí pouze za splnění několika podmínek. První dvě a zároveň ty nejdůležitější jsou vychýlení žáka s hodnotou *T-Wert* od průměru. Ten se uvažuje s dolní hranicí 40 a vrchní 60 bodů. Pokud je studentovo hodnota mimo tento rozsah, program zachytí tuto skutečnost a automaticky provádí veškeré kroky ke zpracování a vizualizaci dat. Naopak při nesplnění podmínky se ve výsledném dokumentu nevyskytne ani zmínka o této funkci programu. Přídavnou podmínkou pro zobrazení grafu je volba typu hodnoceného subjektu v uživatelském rozhraní. Pokud uživatel zvolí jiný subjekt než žáka, program nedokáže vyhodnotit tuto volbu a zároveň je nelogické vyhodnocovat soubor, pokud u každého žáka je krabicový graf stejný vyjma tyrkysového ukazatele odpovědi dotyčného.

4.7 Optimalizace programu

Každý program je nutné konstruovat a mít přítom na paměti časovou náročnost. Vzhledem k povaze vyvíjeného programu, kdy se bude využívat velmi často a zároveň vícekrát za sebou v rychlé sekvenci, je nutné co nejvíce snížit jak náročnost při běhu programu, tak i celkové časové vytížení. K tomu existuje mnoho způsobů při různých operacích, jako například u vyhledávání se uplatňují různé algoritmy, které jsou mnohem efektivnější než ortodoxní přístupy k datům.

V případě tvořeného programu nebylo nutné implementovat žádné speciální algoritmy či postupy, které by zrychlily chod programu. Z důvodu vývoje v prostředí programovacího jazyka R, který je určen výhradně pro práci s objemnými daty a ke statistické analýze, jelikož je jeho generické funkce jsou opatřeny výkonnými algoritmy poskytující vysoký výkon.

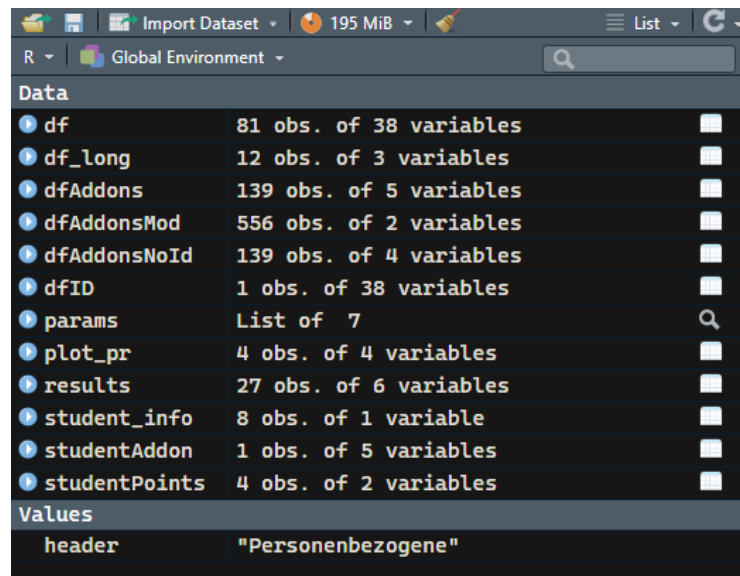
Avšak mohou se naskytnout situace, kdy se může zdát, že program se v jisté sekci kódu zastaví na delší než přívětivou dobu. Kvůli těmto případům jsou vytvářeny nové komunitní knihovny, které obsahují moderní algoritmy, jejichž výkon může být přínosnější než ty v základních funkcích.

Nutné podotknout, že tyto situace nastávají zřídka a pokud ano, tak se jedná o specifické načítání velkého souboru dat, na něž existuje specializovaný přístup. Na druhou stranu se může tento přístup zdát jako neefektivní na soubory s menší velikostí.

V případě programu této práce byla k rychlejšímu načítání dat využita funkce *fread*, která snižuje čas k načtení Excel datasetu se záznamy výsledků žáků. Při testování se snížil čas o necelou sekundu. Toto snížení procentuálně závisí na velikosti dat, proto se může jednat u objemnějších souborů o desítky sekund.

Dalším faktorem ovlivňujícím chod programu je využití paměti programem. Pokud se kód spouští na výkonnější počítačové sestavě s vysokou kapacitou RAM (*Random Access Memory*), není nutné dbát na tento jev. Bohužel ve většině případů nemá člověk k dispozici výkonný počítač a je nutné brát tuto skutečnost v úvahu vzhledem k bezproblémovému chodu a používání programu.

Program si do paměti ukládá pomocné proměnné, které dále využívá při jeho chodu. Z čehož vychází nutnost organizace těchto prvků a také snížení množství na minimum.



Obrázek 4.7 - náhled vytvořených proměnných za celý běh programu, lze vidět v horní části paměťovou náročnost 195 MiB

Některá vývojářská prostředí obsahují tzv. *Garbage collectory*. Starají se o periodické čištění prostoru, který si program vymezil v paměti. V rámci používaného prostředí RStudio však není tato funkce dostupná, a je nutné ručně manipulovat s proměnnými. V našem programu se osvědčila metoda vytváření proměnných pouze pokud je to nutné. V praxi to znamená, že při zvolení jednoho typu hodnoceného subjektu není nutné vytvářet proměnné z dalších typů a zbytečně tak alokovat „prázdné“ místo v paměti pro daný prvek. Využitím podmínek lze dosáhnout tohoto chování, čímž dojde ke zrychlení celého chodu programu a snížení paměťové zátěže, což lze zpozorovat na obrázku 4.7.

Průběh spouštění kódu v rámci tohoto projektu nebyl v jiných částech zpomalován, a proto se nenaskytly jiné možnosti k uplatnění výkonných algoritmů či postupů k efektivnější práci s daty.

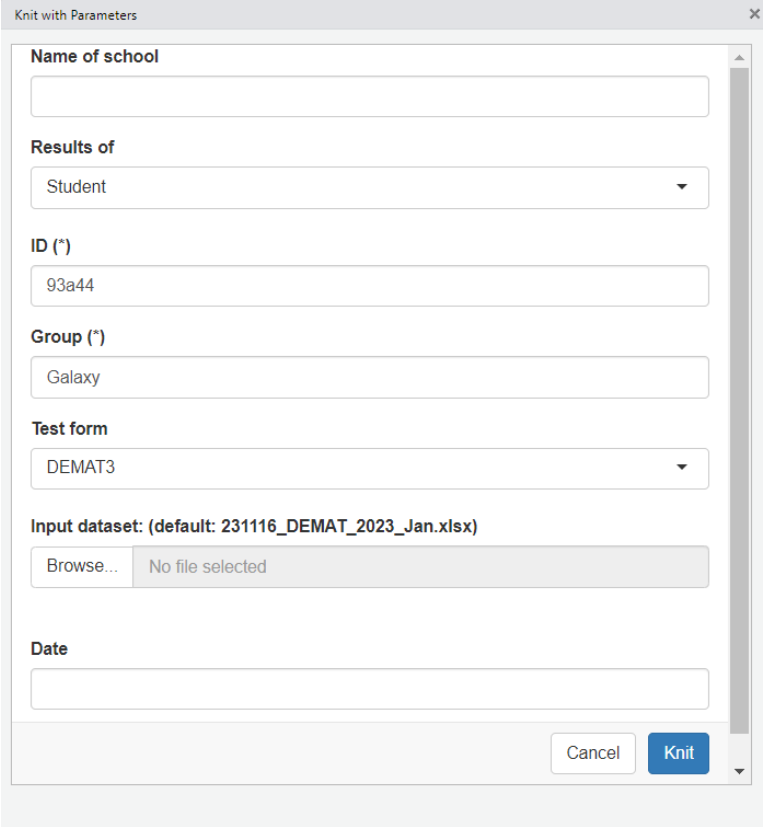
5 Implementace a výsledky

Tato kapitola se bude věnovat popisu implementace uživatelského rozhraní, pomocí kterého může uživatel měnit výchozí parametry užívané při tvorbě reportu. Druhá část se zabývá výsledným vzhledem generovaného dokumentu ve formátu PDF. Dataset není v této části zmíněn, jelikož se jedná o soubor poskytnut externě s vyhrazenými právy čtení a částečné úpravy dat.

5.1 Uživatelské rozhraní

Uživatelské rozhraní (UI) se vygenerovalo s pomocí knihovny *knitr*, která obsahuje specializované funkce k dynamickému vytváření dokumentů v programovacím jazyce R. Celé UI bylo navrženo podle potřeb TU a zároveň v mezích, kterých je schopno dosáhnout s užitou technologií.

UI aplikace (Obrázek 5.1) obsahuje položky označené hvězdou, jejichž vyplnění je zásadní pro chod programu. Naproti tomu se v rozhraní vyskytují i pole bez označení, která není nutné vyplnit bez ohledu na správné chování programu.



The image shows a dialog box titled "Knit with Parameters". It contains the following fields and controls:

- Name of school**: A text input field.
- Results of**: A dropdown menu with "Student" selected.
- ID (*)**: A text input field containing "93a44".
- Group (*)**: A text input field containing "Galaxy".
- Test form**: A dropdown menu with "DEMAT3" selected.
- Input dataset: (default: 231116_DEMAT_2023_Jan.xlsx)**: A file selection area with a "Browse..." button and the text "No file selected".
- Date**: A text input field.
- At the bottom right, there are "Cancel" and "Knit" buttons.

Obrázek 5.1 - Náhled uživatelského rozhraní s jednotlivými poli

Jedná se o vstupní data, jejichž absence se promítne pouze v tabulce s referenčními údaji jako prázdné pole. Další vliv na funkčnost aplikace tato proměnná již nemá.

Prvním polem s názvem *Name of school* je právě takovou položkou. Pokud uživatel toto pole nevyplní, ochudí se o možnost kategorizace reportu. Další pole pojmenované jako *Results of* je vitální částí, která mění samotné fungování programu v oblasti výpočtu a vykreslení jednotlivých částí. Již při prvním spuštění je její výchozí hodnota nastavena na typ *Student*. Zvolením dané možnosti se změní typ hodnoceného subjektu. Výběr jednotlivých typů je zprostředkován pomocí rozbalovací nabídky se třemi možnostmi: student, třída a škola.

Pole *ID* se v případě zvolení typu subjektu vyhodnocení „Student“ nastaví jako povinné. Opatřeno je i doprovodným znakem informujícím o jeho nutnosti vyplnění ve zmíněném případě. V neposlední řadě se pole s názvem *Group* využívá ke specifikaci třídy, která má být vyhodnocena jako celek. Povinnost vyplnění se stahuje na toto pole s identickými podmínkami jako u předešlého, záleží tedy na volbě hodnoceného subjektu.

Další část UI nesoucí název *Test form* určuje formát testu, který daný subjekt podstoupil, nebo si uživatel přeje vyhodnotit. Nabídka se znovu vytvořila jako rozbalovací nabídka nabízející v momentě psaní této práce dva typy testů: DEMAT3 a DEMAT5. Momentálně se na škole s nejvyšším počtem tříd využívá až 7 druhů testů. Po diskuzi a zvážení možností s pracovníky TU se došlo k rozhodnutí, že stačí v této chvíli verze programu umožňující zpracování zmíněných dvou typů testů.

Nejdůležitějším prvkem v UI je pole nesoucí název *Input dataset*. Vybraný soubor tvoří hlavní pilíř funkčnosti celého programu. V této verzi aplikace je podporován formát souboru *.xlsx* nebo *.xls*. Vyplnění pole není povinné, pokud má uživatel přístup k souboru, jehož jméno se nachází nad polem. Pokud tomu tak není, uživatel musí neprodleně vybrat soubor k načtení. K vyhledání a volbě souboru se znovu využila funkčnost knihovny *knitr*, která nabízí integraci se systémem Windows, a jeho příslušným průzkumníkem souborů. Lze tak intuitivně nabídnout uživateli pohodlné prostředí s jehož funkcností je již obeznámen.

Poslední částí UI je pole *Date*. Lze ho využít ke třídění jednotlivých reportů díky propsání do referenční tabulky. Není nutné ho však vyplňovat, pokud tak uživatel nechce učinit. Zůstane-li nevyplněné, program automaticky volí jeho výchozí hodnotu jako aktuální datum systému, ve kterém se spouští.

5.2 Vzhled vytvořeného reportu

Výstup programu ve formě PDF dokumentu se ve své finální podobě vmístil na dvě strany. Obrázek 5.2 znázorňuje první stranu vzorového reportu.

Ergebnisrückmeldung DEMAT3+

Personenbezogene Daten

Identifizier	13025
Klasse	3a
Erhebungszeitraum	duben 2024
Vergleichswerte	—
Klassenstufe	—
Name der Schule	TU Dresden
Jahrgang	—
Testform	DEMAT3+

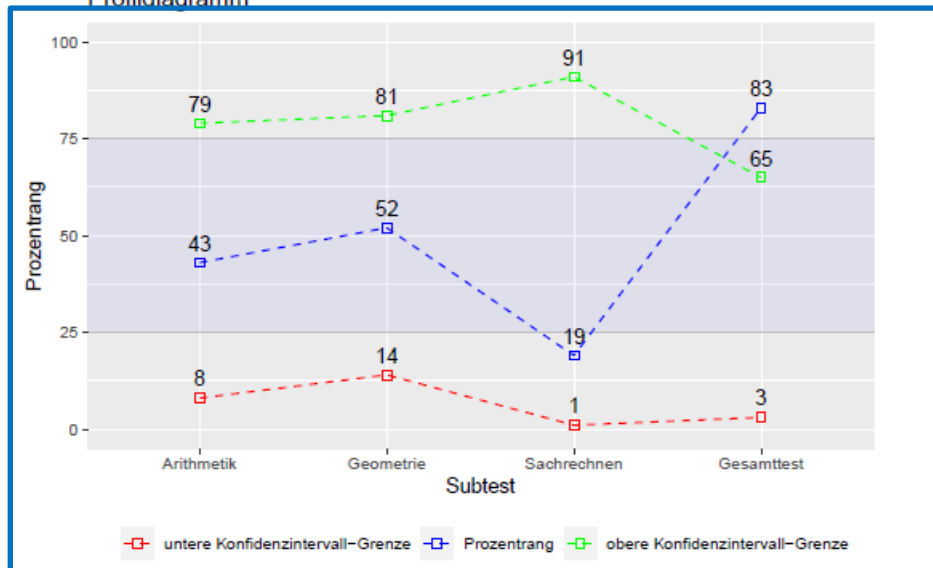
1

Testergebnisse

Aufgabentyp	Rohwerte			
ZS Zahlenstrahl	3			3
AD Addition	3			3
SB Subtraktion	2			2
MP Multiplikation	3			3
SA Sachaufgaben		2		2
SP Spiegelung			3	3
FL Formen legen			2	2
LS Längen schätzen			2	2
LU Längen umrechnen		4		4
Subtest	Arithmetik	Sachaufgaben	Geometrie	Gesamtwert
Rohwert (Summe)	11	6	7	24
Prozentrang	43	19	52	83
T-Wert				59

2

Profildiagramm



3

1

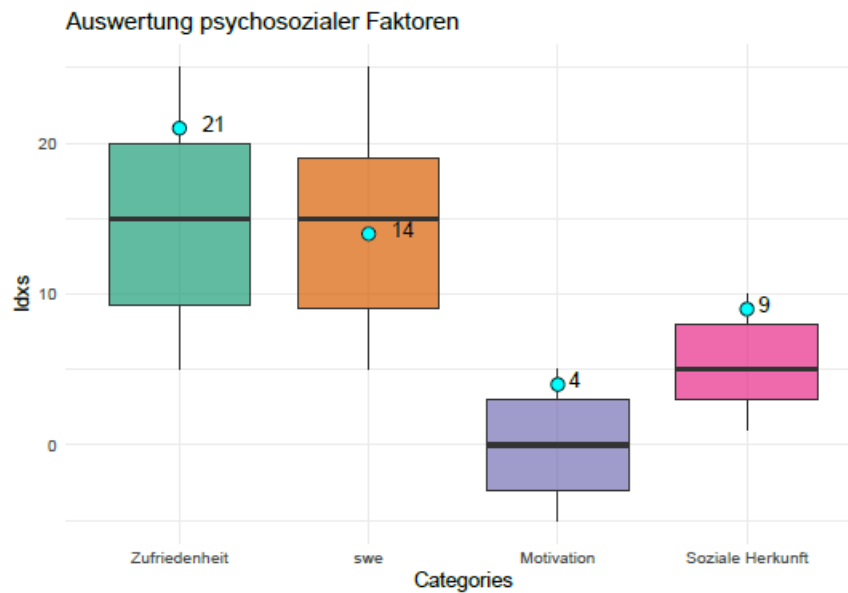
Obrázek 5.2 - první strana dokumentu (1 – referenční údaje subjektu, 2 - detailní ohodnocení jednotlivých disciplín, 3 – percentilové umístění subjektu a interval spolehlivosti)

První označená tabulka obsahuje veškeré referenční údaje o hodnoceném subjektu a jiné podrobnosti, které může uživatel měnit dle libosti.

Tabulka označena číslem dva podává podrobné informace o výsledcích subjektu v testu, včetně shrnutí ve formě bodových průměrů jednotlivých okruhů.

Graf nacházející se v dolní části dokumentu (Obr. 5.2) vyjadřuje percentil, kterého subjekt dosáhl. Je rozdělen na celkem čtyři části podle tematického okruhu testu. Individuální body jsou spojeny ke zřetelnějšímu znázornění hodnot a jejich změn mezi skupinami.

Poslední prvek dokumentu nacházejícího se na jeho druhé straně lze zahlédnout na obrázku 5.3. Představuje odpovědi na otázky v anketě podané žákům v průběhu roku týkající se jejich sociálního zázemí a mentálního zdraví.



2

Obrázek 5.3 - krabicové grafy reprezentující bodové hodnocení otázek v sociologické anketě žákem

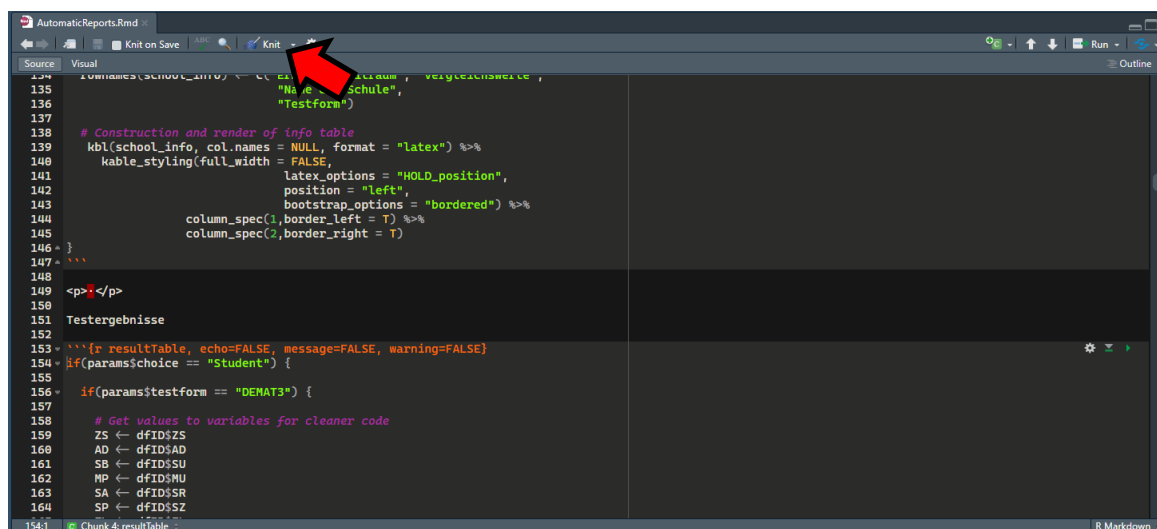
6 Uživatelská dokumentace

V této kapitole je popsán postup zprovoznění aplikace a případná řešení nesprávného chování.

Nejdůležitějším prvkem ke spuštění aplikace je mít na svém zařízení nainstalovaný programovací jazyk R a vývojové prostředí RStudio a jejich nejnovější verze (R: 4.4.0¹, RStudio: 2023.12.1+402²).

Důležitým krokem ke zprovoznění programu je umístění datasetů ve formě Excel tabulek, které jsou doloženy v Seznamu příloh. V případě této aplikace stačí pouze umístit dané soubory do stejné složky jako skript programu. Alternativně lze soubor dohledat pomocí průzkumníka souborů.

Po otevření kódu programu se uživatel ocitne v editoru RStudio. Ke spuštění aplikace a přesunutí se do uživatelského rozhraní ke komunikaci s aplikací stačí v horní části editoru nalézt tlačítko *Knit* s obrázkem klubka nitě (viz Obr. 6.1.). Pomocí rozbalovací nabídky zvolte *Knit with parameters...*



Obrázek 6.1 - vyznačení tlačítka ke spuštění aplikace ve vývojovém prostředí RStudio

¹ <https://cran.r-project.org/bin/windows/base/>

² <https://posit.co/download/rstudio-desktop/>

Po vykonání akcí v předešlém odstavci se otevře okno uživatelského rozhraní (viz Kap. 5.1). Při volení a zadávání jednotlivých identifikátorů či jiných možností je nutno dbát na správnost výběru formy testu a také samotného identifikátoru žáka při zvolení příslušné možnosti typu hodnoceného subjektu.

Pokud se PDF soubor nevytvoří (základní adresář je ve složce se skriptem) či jeho hodnoty nejsou správné, pak je nutné ověřit správnost veškerých údajů včetně jejich srovnání s tabulkou přímo v Excelu odkud data program získává.

Také je nutné brát na vědomí, že některé datasey obsahují data k jednotlivým formám testu, proto je důležité ověřit v Excelu jejich přítomnost. Dále se kvůli změnám v systému hodnocení měnila struktura jednotlivých datasetů, a proto pro některé nemusí být report celistvý a mohou se v něm vyskytovat nulové hodnoty. Tento problém se však v rámci testování aplikace vyjímá jako neškodný, jelikož veškeré dodané datasey byly jednotlivě otestovány na částech aplikace. Příkladem tohoto chování je použití jednoho datasetu a správného navolení výběrových parametrů, avšak ve výsledném dokumentu mohou být v tabulce s výsledky samé nuly, či neexistující intervaly a body v grafu.

Daný problém nelze momentálně vyřešit kvůli nejednotné struktuře dodaných datasetů. Avšak při testování aplikace se nevyskytly u jednotlivých částí žádné problémy se zpracováním dat.

7 Diskuse

Výstupem této práce je plně funkční aplikace umožňující automatizovaně vytvářet dynamické dokumenty obsahující výsledky ze standardizovaných testů DEMAT. V rámci této bakalářské práce byly úspěšně splněny veškeré požadavky zadané Technickou univerzitou v Drážďanech. Program je dále připraven k možnému navazujícímu vývoji skrze komunikaci s pověřenými osobami na TU. Samotná operace s aplikací je intuitivní a kódová část se může libovolně upravovat podle představ IT experta.

Program lze spustit na jakémkoliv zařízení, které podporuje instalaci vývojového prostředí RStudio a programovací jazyk R. Momentálně však uvedené technologie podporují pouze instalaci na operační systémy Windows 8 a vyšší, Linux (nebo jeho další UNIX ekvivalenty) a macOS. Rozšíření nebo případné spouštění na jiných systémech závisí na vývojovém týmu daných produktů. Lze konstatovat, že v budoucnu se může vyskytnout stand-alone verze vytvořeného programu v podobě spustitelného balíčku s příponou .exe. Pokud by se tak stalo, byla by odstraněna menší prodleva v tvorbě jednotlivých reportů.

Aplikace se dočkala nových vylepšení a funkcí, které nebyly na počátku tohoto projektu definované a ani se o nich neuvažovalo, kupříkladu způsob prezentace výsledků ankety, týkající se psycho- a socioekonomické situace individuálního studenta. Naproti tomu má program prostor pro vylepšení a nové funkce. Jedním z důležitých dodatků pro přehlednost a čitelnost souborů produkovaných tímto programem je úprava grafického zpracování. Dalším vylepšením, které by se mohlo pozitivně projevit na časové náročnosti kódu, by bylo vytvoření spustitelného balíčku, díky němuž se eliminuje nutnost instalace programovacího jazyka a jeho vývojového prostředí.

Program lze také využít v jiných oborech, například ve zdravotnictví. V dnešní době se společnost často potýká s problémem přebytku zdravotnických dat pacientů s nutností zpracování do formy, která je užitečná pro doktory a zdravotnický personál. Program je možno jednoduše přenastavit na načítání a zpracovávání datasetů obsahující data pacientů. Po stanovení potřebných nezpracovaných dat nemocnicí může program produkovat potřebné reporty přibližující pacientův stav.

8 Závěr

Cílem práce bylo vytvoření programu, který odpovídá veškerým požadavkům TU v Drážďanech. Jeho hlavním úkolem je automatizace generování reportu se statistickými výsledky o datech pocházejících z matematických testů DEMAT. V pozdější fázi jeho budování bylo také vyzdviženo zohlednění a analýza vnějších vlivů na studentovo výsledky. Podoba vyobrazení byla dohodnuta na schůzi s jedním z pracovníků na TU v Drážďanech. Výsledkem této diskuse se stal graf typu boxplot.

Program byl vytvořen za pomoci programovacího jazyka R, který je navržen pro programy zacházejícími se statistickými výpočty a komplexními datovými strukturami, jehož výstupem jsou dynamické dokumenty, které lze vizuálně přizpůsobovat. Dále bylo využito nativní vývojové prostředí RStudio a několik pomocných knihoven, jejichž funkce usnadnily práci s mnohotvárnými daty. V průběhu jeho konstrukce se aktivně testoval zprvu na umělých, a v dalších verzích na reálných datech z několika škol a ročníků, které spadaly pod projekt Unischule Dresden. V rámci testování a požadavků bylo nutné uzpůsobit program ke zpracování dat ze dvou různých forem testů, jejichž datové struktury se patřičně lišily.

Dokument se dočkal v každé jeho verzi nového vzhledu včetně nových funkcí. Data jsou usazena ve dvou základních tabulkách, kdy první obsahuje veškeré referenční údaje hodnoceného subjektu a druhá, hlavní, prezentuje podrobné výsledky podstoupeného testu.

Funkce programu, která si klade za cíl přiblížit a ulehčit analýzu vlivu vnějších faktorů na akademické výsledky, se testovala na umělých datech, které byly poskytnuty TU. Důvodem užití umělých dat byla anketa, která se teprve nedávno zahrnula do výukového systému na školách akademické obce v Drážďanech. Proto nebyl dostupný dataset s novými daty. Každopádně bylo řečeno vedením TU, že dataset si udrží stejnou datovou strukturu a nebude se výrazně měnit jeho podoba, jejíž změna by měla za následek nesprávný chod programu.

Výsledek této práce se již přes rok aktivně využívá k vyhodnocení uskutečněných kol testování.

Seznam použité literatury

- [1] BROUSSARD, Sheri Coates a M. E. Betsy GARRISON. The Relationship Between Classroom Motivation and Academic Achievement in Elementary-School-Aged Children. *Family and Consumer Sciences Research Journal* [online]. 2004, **33**(2), 106–120. ISSN 1077-727X, 1552-3934. Dostupné z: doi:10.1177/1077727X04269573
- [2] LANGNER, Anke a Maxi HESS. Die Universitätsschule Dresden. *WE_OS Jahrbuch* [online]. 2020, 11-36 Seiten. Dostupné z: doi:10.4119/WE_OS-3340
- [3] UNIVERSITÄTSSCHULE DRESDEN. Die Design-Thinking-Räume. *Padlet* [online]. [vid. 2024-05-13]. Dostupné z: <https://padlet.com/universitaetsschule/der-digitale-rundgang-der-universitaetsschule-dresden-n3d1mcymn38osxcd/wish/601753891>
- [4] MULTIP.CZ. *ŠKOLNÍ TRÍDY* [online]. [vid. 2024-05-13]. Dostupné z: <https://www.multip.cz/nabizime-vybaveni-pro-skolni-tridy>
- [5] GRÄF, Christine, Isabell HOFFMANN, Christiane DIEFENBACH, Jochem KÖNIG, Martina F. SCHMIDT, Kathleen SCHNICK-VOLLMER, Michael HUSS a Michael S. URSCHITZ. Mental health problems and school performance in first graders: results of the prospective cohort study ikidS. *European Child & Adolescent Psychiatry* [online]. 2019, **28**(10), 1341–1352. ISSN 1018-8827, 1435-165X. Dostupné z: doi:10.1007/s00787-019-01296-7
- [6] SUUTELA, Maria, Päivi J. MIETTINEN, Silja KOSOLA, Ossi RAHKONEN, Tero VARIMO, Annika TARKKANEN, Matti HERO a Taneli RAIVIO. Timing of puberty and school performance: A population-based study. *Frontiers in Endocrinology* [online]. 2022, **13**, 936005. ISSN 1664-2392. Dostupné z: doi:10.3389/fendo.2022.936005
- [7] VAN ZWIETEN, Anita, Armando TEIXEIRA-PINTO, Suncica LAH, Natasha NASSAR, Jonathan C. CRAIG a Germaine WONG. Socioeconomic Status During Childhood and Academic Achievement in Secondary School. *Academic Pediatrics* [online]. 2021, **21**(5), 838–848. ISSN 18762859. Dostupné z: doi:10.1016/j.acap.2020.10.013
- [8] CHU, Tianshu, Xin LIU, Shigemi TAKAYANAGI, Tomoko MATSUSHITA a Hiro KISHIMOTO. Association between mental health and academic performance among university undergraduates: The interacting role of lifestyle behaviors. *International Journal of Methods in Psychiatric Research* [online]. 2023, **32**(1), e1938. ISSN 1049-8931, 1557-0657. Dostupné z: doi:10.1002/mpr.1938
- [9] DESTIN, Mesmin, Paul HANSELMAN, Jenny BUONTEMPO, Elizabeth TIPTON a David S. YEAGER. Do Student Mindsets Differ by Socioeconomic Status and Explain Disparities in Academic Achievement in the United States? *AERA Open* [online]. 2019, **5**(3), 233285841985770. ISSN 2332-8584, 2332-8584. Dostupné z: doi:10.1177/2332858419857706
- [10] SAGIROGLU, Seref a Duygu SINANC. Big data: A review. In: *2013 International Conference on Collaboration Technologies and Systems (CTS): 2013 International Conference on Collaboration Technologies and Systems (CTS)* [online]. San Diego, CA, USA: IEEE, 2013, s. 42–47 [vid. 2024-03-08]. ISBN 978-1-4673-6404-1. Dostupné z: doi:10.1109/CTS.2013.6567202

- [11] GARCÍA, Salvador, Sergio RAMÍREZ-GALLEGO, Julián LUENGO, José Manuel BENÍTEZ a Francisco HERRERA. Big data preprocessing: methods and prospects. *Big Data Analytics* [online]. 2016, **1**(1), 9. ISSN 2058-6345. Dostupné z: doi:10.1186/s41044-016-0014-0
- [12] MISHRA, Puneet, Alessandra BIANCOLILLO, Jean Michel ROGER, Federico MARINI a Douglas N. RUTLEDGE. New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry* [online]. 2020, **132**, 116045. ISSN 01659936. Dostupné z: doi:10.1016/j.trac.2020.116045
- [13] LUENGO, Julián, Diego GARCÍA-GIL, Sergio RAMÍREZ-GALLEGO, Salvador GARCÍA a Francisco HERRERA. *Big Data Preprocessing: Enabling Smart Data* [online]. Cham: Springer International Publishing, 2020 [vid. 2024-03-04]. ISBN 978-3-030-39104-1. Dostupné z: doi:10.1007/978-3-030-39105-8
- [14] GIORGI, Federico M., Carmine CERAOLO a Daniele MERCATELLI. The R Language: An Engine for Bioinformatics and Data Science. *Life* [online]. 2022, **12**(5), 648. ISSN 2075-1729. Dostupné z: doi:10.3390/life12050648
- [15] Understanding Boxplots. *KDnuggets* [online]. [vid. 2024-05-14]. Dostupné z: <https://www.kdnuggets.com/understanding-boxplots>

Seznam obrázků

Obrázek 2.1 – místnost pro žáky určena k vývoji design-thinking [3]	7
Obrázek 2.2 - názorná ukázka třídy v České republice [4]	8
Obrázek 2.3 – tabulka s odhadem efektu mentálních problémů na výsledky [3]	11
Obrázek 2.4 - tabulka studie zaměřené na efekt nízkého SES na z-skóre [7]	12
Obrázek 2.5 - graf rozdělení žáků do skupin podle dosažených bodů v K6 [6]	14
Obrázek 2.6 - vztah mezi úrovní SES a statickým typem myšlení [9].....	15
Obrázek 4.1 - struktura listů v datasetu	21
Obrázek 4.2 - tabulka s referenčními údaji hodnoceného subjektu	26
Obrázek 4.3 - tabulka s podrobnými výsledky testu	28
Obrázek 4.4 - náhled grafu znázorňujícího percentil a interval spolehlivosti	30
Obrázek 4.5 - ukázka krabicového grafu včetně popisu jeho částí [15].....	32
Obrázek 4.6 - názorná ukázka boxplotu v reportu	33
Obrázek 4.7 - náhled vytvořených proměnných za celý běh programu	36
Obrázek 5.1 - Náhled uživatelského rozhraní s jednotlivými poli (autor)	37
Obrázek 5.2 - první strana vygenerovaného dokumentu.....	39
Obrázek 5.3 - bodové hodnocení otázek v sociologické anketě žákem	40
Obrázek 6.1 - vyznačení tlačítka ke spuštění aplikace ve vývojovém prostředí	41