

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

OPTIMALIZÁCIA JEDNOTKOVÝCH CIEN STAVEBNÝCH
PRÁC A MATERIÁLOV PRE BENCHMARKING
ŽELEZNIČNÝCH INVESTIČNÝCH PROJEKTOV

DIPLOMOVÁ PRÁCA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**OPTIMALIZÁCIA JEDNOTKOVÝCH CIEN STAVEBNÝCH
PRÁC A MATERIÁLOV PRE BENCHMARKING
ŽELEZNIČNÝCH INVESTIČNÝCH PROJEKTOV**

DIPLOMOVÁ PRÁCA

Študijný program: Ekonomicko-finančná matematika a modelovanie
Študijný odbor: Matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: Mgr. Matúš Rako



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Tomáš Starek
Študijný program: ekonomicko-finančná matematika a modelovanie
(Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Optimalizácia jednotkových cien stavebných prác a materiálov
pre benchmarking železničných investičných projektov
*Optimization of unit prices of construction works and materials for
benchmarking of railway investment projects*

Anotácia: Vytvorenie cenovej databázy jednotkových cien stavebných prác a materiálov
z už realizovaných železničných investičných projektov, odhadovanie
a optimalizácia týchto jednotkových cien matematicko-štatistickými metódami
pre potreby benchmarkingu železničných investičných projektov Útvorom
hodnoty za peniaze a porovnanie s existujúcimi spôsobmi odhadovania týchto
jednotkových cien.

Vedúci: Mgr. Matúš Rako
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: doc. Mgr. Igor Melicherčík, PhD.
Dátum zadania: 11.01.2022

Dátum schválenia: 14.01.2022

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Podakovanie Prvotne sa chcem poďakovať svojmu školiťovi Mgr. Matúšovi Rakovi za jeho veľkú dávku trpezlivosti, za prejavení čas a za množstvo skvelých rád s ktorými ma sprevádzal počas písania predloženej diplomovej práce. Taktiež mu ďakujem za množstvo hodín spoločných podnetných diskusií, z ktorých si odnášam neoceniteľné vedomosti do ďalšieho života. Chcem sa taktiež poďakovať svojej rodine za to, že ma finančne podporovali počas môjho štúdia a boli tu vždy pre mňa, keď som to potreboval. Ďakujem aj svojej milujúcej priateľke, ktorá ma podporovala a motivovala počas písania predloženej práce. V neposlednom rade sa chcem poďakovať každému, kto si nájde čas prečítať túto diplomovú prácu. Vaša snaha a záujem o získavanie nových poznatkov svedčí o vášnivej túžbe po vzdelávaní, ktorá nás všetkých spája. Ďakujem vám za to, že ste súčasťou tejto vzdelávacej cesty.

Abstrakt v štátnom jazyku

STAREK, Tomáš: Optimalizácia jednotkových cien stavebných prác a materiálov pre benchmarking železničných investičných projektov [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Matúš Rako, Bratislava, 2023, 74 s.

Diplomová práca je zameraná na navrhnutie vlastného modelu odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov pre potreby Oddelenia dopravných projektov Útvaru hodnoty za peniaze pri benchmarkovaní predpokladanej hodnoty týchto projektov ako alternatívu oproti súčasnému spôsobu benchmarkovania týchto projektov pomocou cenových databáz českého Štátneho fondu dopravnej infraštruktúry. Vychádza z dôkladne spracovaných a matematicko-štatistickými metódami analyzovaných železničných zmluvných rozpočtov, pričom predstavuje vlastný mechanizmus štandardizácie váh týchto rozpočtov, vlastný systém prevedenia jednotkových cien na spoločnú cenovú úroveň, ako aj vlastnú metódu odhadovania jednotkových cien. Záverečná časť práce popisuje validáciu navrhnutého modelu na dátach simulujúcich reálne podmienky v praxi a to aj v konfrontácii s ocenenými výkazmi výmer železničných investičných projektov od projektantov, súčasným spôsobom benchmarkovania železničných investičných projektov Oddelením dopravných projektov Útvaru hodnoty za peniaze, ako aj s jednotkovými cenami stavebných prác a materiálov komerčnej cenovej databázy CENEKON. Ku koncu predloženej práce demonštruje stochastická simulácia vysoký potenciál presnejšieho odhadovania jednotkových cien navrhnutým modelom pri ďalšom rozširovaní cenovej databázy. Práca má ambíciu ponúknuť model odhadovania jednotkových cien stavebných prác a materiálov nielen Oddeleniu dopravných projektov Útvaru hodnoty za peniaze, ale aj ďalším štátnym inštitúciám Slovenskej republiky pracujúcimi s rozpočtami železničných investičných projektov, ako aj možnosť adaptácie modelu na širšie spektrum investičných projektov.

Kľúčové slová: benchmarking, cenová databáza, jednotkové ceny, odhadovanie, železničné projekty

Abstract

STAREK, Tomáš: Optimization of unit prices of construction works and materials for benchmarking of railway investment projects [Master Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Matúš Rako, Bratislava, 2023, 74 p.

This master's thesis is focused on the development of our own model for estimating unit costs of construction work and materials for railway investment projects, tailored to the needs of the Department of Transportation Projects under the Value for Money Division. The model is designed as an alternative to the current method of benchmarking these projects using price databases of the czech State Fund for Transport Infrastructure. The thesis builds on thoroughly processed and mathematically-statistically analysed railway contract budgets, proposing a unique mechanism for standardizing the weights of these budgets, our own system for converting unit prices to a common price level, as well as a novel method for estimating unit prices. The final part of the thesis describes the validation of the proposed model using data that simulate real-world conditions in practice, comparing it with the priced bills of quantities for railway investment projects from designers, the current method of benchmarking railway investment projects by the Department of Transportation Projects under the Value for Money Division, as well as the unit prices of construction work and materials in the commercial price database CENEKON. Towards the end of the thesis, a stochastic simulation demonstrates the high potential for a more accurate unit cost estimation using the proposed model as the price database expands. The thesis aims to offer a model for estimating unit costs of construction work and materials not only to the Department of Transportation Projects under the Value for Money Division but also to other state institutions in the Slovak Republic working with budgets for railway investment projects, as well as the possibility of adapting the model to a broader spectrum of investment projects.

Keywords: benchmarking, price database, unit prices, estimation, railway projects

Obsah

Zoznam obrázkov	9
Zoznam tabuliek	10
Zoznam použitých skratiek	11
Úvod	13
1 Problematika odhadovania nákladov železničných investičných projektov	15
1.1 Projektová príprava a odhadovanie nákladov	15
1.2 Rozpočty železničných investičných projektov	17
2 Potrebne poznatky zo štatistiky	24
2.1 Metódy na identifikáciu nestability v dátach	24
2.2 Metódy na identifikáciu vychýlenosti dát	25
2.3 Metódy na identifikáciu outlierov v dátach	25
2.4 Metódy na odhadovanie bodovej hodnoty	27
2.5 Metódy na validovanie odhadu modelom	28
2.6 Metódy na meranie presnosti odhadu modelom	30
3 Metodika tvorby cenovej databázy a analýzy dát	32
3.1 Tvorba cenovej databázy	33
3.2 Čistenie dát v cenovej databáze	36
3.3 Štandardizácia váh projektov	37
3.4 Zjednotenie cenovej úrovne	38
3.5 Identifikácia nestabilných položiek v dátach	39
3.6 Identifikácia outlierov v dátach	41
3.7 Bodové odhadovanie jednotkových cien	42
3.8 Porovnávanie metód odhadovania jednotkových cien	44
3.9 Testovanie modelu na dátach simulujúcich reálne podmienky v praxi	46

4	Štatistická analýza a výsledky	49
4.1	Analýza systémov indexácie	49
4.2	Analýza charakteru dát	51
4.3	Analýza vychýlenosti dát	54
4.4	Analýza metód identifikujúcich outliere	56
4.5	Analýza existencie objemnej zľavy	57
4.6	Výsledky krížovej validácie	60
4.6.1	Kalibrovanie metódy optimálny percentil	60
4.6.2	Porovnanie metód odhadovania jednotkových cien	62
4.7	Model odhadovania jednotkových cien	62
4.8	Porovnanie s inými cenovými databázami	63
4.9	Analýza robustnosti cenovej databázy	66
	Záver	68
	Zoznam použitej literatúry	71
	A Prvá príloha	
	B Druhá príloha	
	C Tretia príloha	
	D Štvrtá príloha	

Zoznam obrázkov

4.1	Početnosť unikátnych záznamov položiek v trénovacej databáze	52
4.2	Histogram normalizovaných jednotkových cien v trénovacej databáze	55
4.3	Korelácia množstva a jednotkových cien položiek v trénovacej databáze . .	59
4.4	Hodnota RMSE/MAE v závislosti od počtu projektov v databáze	66
A.1	Plánovanie a príprava investičných projektov na Slovensku	

Zoznam tabuliek

1.1	Roztriedenie objektov podľa profesijných odborov	18
1.2	Príklad nesynchronizácie kódov položiek	19
1.3	Priradenie indexov ku profesijným odborom podľa systému kódov odborov	21
3.1	Zoznam kľúčových slov pre systém názvov objektov a im prináležiace indexy	39
4.1	Výsledky kategorizovaného úhrnného výberového variančného koeficientu pre rôzne systémy indexácie	50
4.2	Rozdelenie kódov profesijných odborov v trénovacej databáze	53
4.3	Početnosť položiek podľa ich vychýlenosti	56
4.4	Výsledky porovnania metód identifikujúcich outliere	56
4.5	Roztriedenie položiek podľa rozdielu váženého priemeru od aritmetického priemeru	58
4.6	Porovnanie metód odhadovania jednotkových cien krížovou validáciou . . .	63
4.7	Hlavné výsledky benchmarku testovacej dátovej sady	64
4.8	Výsledky benchmarkovania podľa vybraných profesijných odborov	65
D.1	Výsledky benchmarku pre model – po projektoch	
D.2	Výsledky benchmarku pre SFDI – po projektoch	
D.3	Výsledky benchmarku pre CENEKON – po projektoch	
D.4	Výsledky benchmarku pre model – po profesijných odboroch	
D.5	Výsledky benchmarku pre SFDI – po profesijných odboroch	
D.6	Výsledky benchmarku pre CENEKON – po profesijných odboroch	

Zoznam použitých skratiek

a i. – a iné

angl. – anglicky

APE – absolútna percentuálna odchýlka (angl. *absolute percentage error*)

atď. – a tak ďalej

č. – číslo

DPH – daň z pridanej hodnoty

DRS – dokumentácia na realizáciu stavby

DSP – dokumentácia na stavebné povolenie

DSPRS – dokumentácia na stavebné povolenie v podrobnostiach na realizáciu stavby

DÚR – dokumentácia na územné rozhodnutie

et al. – a ostatní

FIDIC – Medzinárodná federácia konzultačných inžinierov

FMFI UK – Fakulta matematiky, fyziky a informatiky Univerzity Komenského
v Bratislave

GSM-R – globálny systém mobilnej komunikácie pre železnice

IFP – Inštitút finančnej politiky

IQR – medzikvartilový rozsah

MAE – stredná absolútna odchýlka (angl. *mean absolute error*)

MAPE – stredná absolútna percentuálna odchýlka (angl. *mean absolute percentage error*)

MC – mediánový pár hodnôt (angl. *medcouple*)

MD SR – Ministerstvo dopravy Slovenskej republiky

MF SR – Ministerstvo financií Slovenskej republiky

mil. – milión

mld. – miliarda

MP – metodický pokyn

napr. – napríklad

obr. – obrázok

OTSKP – Odborový triednik stavebných konštrukcií a prác

PHZ – predpokladá hodnota zákazky

PS – prevádzkový súbor

Q1 – dolný kvartil

Q3 – horný kvartil

resp. – respektíve

RMSE – druhá odmocnina strednej kvadratickej odchýlky (angl. *root mean squared error*)

SFDI – Štátny fond dopravnej infraštruktúry

SO – stavebný objekt

str. – strana

ŠR – štúdia realizovateľnosti

št. hr. – štátna hranica

ŠÚ SR – Štatistický úrad Slovenskej republiky

t. j. – to jest

tab. – tabuľka

tis. – tisíc

TSKP – Triednik stavebných konštrukcií a prác

tzv. – takzvaný

ÚHP – Útvar hodnoty za peniaze

ÚOŽI – Zborník pre údržbu a opravy železničnej infraštruktúry

VB – výpravná budova

wMAPE – vážená stredná absolútna percentuálna odchýlka (angl. *weighted mean absolute percentage error*)

Z. z. – Zbierka zákonov

ŽM – železničný most

ŽSR – Železnice Slovenskej republiky

Úvod

Je zvláštne si predstaviť, že by sme pri svojom bežnom nákupe nejakého tovaru, napríklad potravín, ani orientačne nevedeli, koľko by jednotlivé položky mali stáť a konečnú cenu nákupu by sme zistili až pri pokladni, keď je už potrebné zaplatiť. Bez nejakej orientačnej predstavy o cenách nakupujúcich položiek by sme nevedeli, či vôbec máme dostatok peňazí na daný nákup a či cena zodpovedá tej hodnote, akú pre nás ten tovar má. Podobný problém však štát rieši neustále, keď niekde investuje, napríklad do obnovy a modernizácie železničnej infraštruktúry. Cenu za projekt sa dozvie až po skončení verejného obstarávania. Aby štát mal aspoň odhad hodnoty zákazky, pomáhajú mu s tým viaceré štátne inštitúcie, akou je napríklad aj Útvar hodnoty za peniaze spadajúci pod Ministerstvo financií Slovenskej republiky.

Útvar hodnoty za peniaze vznikol v roku 2016 s úlohou „*zvyšovať hodnotu, ktorú za svoje peniaze verejnosť dostáva. Cieľom je posilniť orientáciu na výsledky pri rozhodovaní o verejných politikách, reguláciách, investíciách a prevádzke štátu. Podstatou hodnoty za peniaze je posúdiť, či peniaze daňovníkov budú skutočne vynaložené najlepšie, ako je možné pre dosiahnutie stanoveného cieľa.*“[17] Jednou z mnohých metód, ktorú tento analytický útvar používa na dosiahnutie svojho cieľa, je benchmarking. Benchmarking je proces porovnávania jednotkových cien nielen stavebných prác a materiálov s cieľom zistiť, či je cena adekvátne hodnote danej položky, a to porovnaním s nejakými referenčnými cenami. Útvar hodnoty za peniaze ale v čase písania predloženej práce nedisponoval vlastnou cenovou databázou pre železničné investičné projekty založenej na dátach zo slovenského prostredia, bol tak odkázaný len na zahraničné alebo komerčné cenové databázy[23].

Tak, ako bežný človek vie lepšie hospodáriť, ak má vopred predstavu, koľko ho nákup v obchode bude stáť, tak isto to platí aj pre štát. Čím reálnejšie vie štát odhadovať ceny investičných projektov, tým flexibilnejší vie byť so svojimi financiami. Keďže štát na investície používa peniaze nás všetkých, malo by nám všetkým záležať, aby verejné financie boli využité čo najefektívnejšie. Cieľom predloženej diplomovej práce bolo preto navrhnúť vlastný model odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektoch pre potreby benchmarkovania Oddelením dopravných projektov Útvaru hodnoty za peniaze.

Tento cieľ sme dosiahli viacerými krokmi. Pre potreby predloženej práce sme najprv vytvorili databázu jednotkových cien stavebných prác a materiálov zo zmluvných rozpočtov už realizovaných železničných investičných projektov. Následne sme zostrojili vlastný mechanizmus štandardizácie váh projektov, nakoľko každý projekt poskytoval hodnotovo rovnocennú informáciu o jednotkových cenách na trhu. Navrhli sme vlastný systém prevedenia jednotkových cien na spoločnú cenovú úroveň, aby sa dali jednotkové ceny z jednotlivých rozpočtov navzájom porovnávať. Predložili sme aj návrh vlastnej metódy odhadovania jednotkových cien. Pomocou matematicko-štatistických metód široko popísaných v literatúre napríklad v [7, 9, 11, 12, 18, 21, 26, 34] a dôkladnej analýze dát sme ešte nastavili hyperparametre nášho modelu a optimalizovali odhadovanie jednotkových cien stavebných prác a materiálov železničných investičných projektov.

Venovali sme sa aj aplikovaniu nášho navrhnutého modelu na dáta simulujúce reálne podmienky v praxi a to aj v konfrontácii s ocenenými výkazmi výmer železničných investičných projektov od projektantov, súčasným spôsobom benchmarkovania jednotkových cien stavebných prác a materiálov železničných investičných projektov Oddelením dopravných projektov Útvaru hodnoty za peniaze pomocou cenových databáz českého Štátneho fondu dopravnej infraštruktúry, ako aj s jednotkovými cenami stavebných prác a materiálov najpoužívanejšou komerčnou cenovou databázou na slovenskom stavebnom trhu CENEKON. Stochastickým modelovaním sme poukázali taktiež na potenciál odhadovania jednotkových cien našim modelom pri budúcom rozširovaní cenovej databázy o nové rozpočty železničných investičných projektov.

Našou snahou bolo navrhnutím modelu odhadovania jednotkových cien stavebných prác a materiálov pomôcť Oddeleniu dopravných projektov Útvaru hodnoty za peniaze efektívnejšie benchmarkovať predpokladané hodnoty železničných investičných projektov. Zároveň sme chceli ponúknuť model, ktorý by mohol byť nápomocný aj ďalším štátnym inštitúciám Slovenskej republiky pracujúcich s týmito typmi projektov pre ich lepšie odhadovanie jednotkových cien. V neposlednom rade sme chceli našou prácou priniesť nové poznatky do oblasti vyvíjania cenových databáz a benchmarkovania jednotkových cien a umožniť tak náš model v budúcnosti adaptovať aj na širšie spektrum investičných projektov, napríklad na investičné projekty cestnej infraštruktúry či pozemných stavieb.

1 Problematika odhadovania nákladov železničných investičných projektov

Nevyhnutnou súčasťou každého investičného projektu je korektne odhadnutá predpokladaná hodnota zákazky (PHZ). Odhad PHZ má vplyv napr. na zaistenie financovania projektu, na alokovanie eurofondov, či efektívnosť hospodárenia s verejnými financiami.

V prvej kapitole našej práce sme preto predstavili súčasný proces odhadovania nákladov investičných projektov na Slovensku. Vychádzali sme voľne najmä zo záverečnej správy [13]. V predloženej práci sme nepopisovali podrobne celý proces projektovej prípravy investičných projektov, nakoľko sme sa zamerali iba na časť procesu projektovej prípravy týkajúcej sa samotného odhadovania nákladov. Druhú časť kapitoly sme venovali už samotnému popisu rozpočtov železničných investičných projektov, či už predpokladaných alebo zmluvných, a aké problémy Útvar hodnoty za peniaze (ÚHP) musí pri práci s nimi riešiť. Vychádzali sme voľne najmä z článku [5] a osobnej konzultácie s ÚHP[23].

1.1 Projektová príprava a odhadovanie nákladov

Proces projektovej prípravy investičných projektov a odhadovania ich nákladov je v slovenských reáliách pomerne zložitý. V rámci plánovania železničných investičných projektov prebieha niekoľko etáp projektovej prípravy. V prvej etape sa formuje zámer projektu, ktorý obsahuje hrubý odhad nákladov na jeho realizáciu. Pre projekty s hodnotou nad 40 mil. € vrátane dane z pridanej hodnoty (DPH) je vyžadované nariadením vlády Slovenskej republiky č. 174/2019 Z. z.[19] vypracovať štúdiu realizovateľnosti (ŠR), ktorú následne hodnotí aj ÚHP. Po zámere projektu nasleduje vypracovanie dokumentácie na územné rozhodnutie (DÚR) a pokračuje sa vypracovaním dokumentácie na stavebné povolenie (DSP). V týchto etapách sa odhad nákladov spresňuje. Pre projekty s hodnotou nad 20 mil. € s DPH sú odhady nákladov v rámci DÚR a DSP v zmysle zákona č. 254/1998 Z. z.[37] podrobené ešte kontrole štátnej expertízy, za ktorú zodpovedá Ministerstvo dopravy Slovenskej republiky (MD SR). Po týchto etapách sa prechádza do poslednej etapy projektovej prípravy, ktorou je dokumentácia na realizáciu stavby (DRS). V tejto etape sa odhad nákladov stanovuje už na úrovni výkazov výmer, tzv. červená kniha podľa zmluvných štandardov Medzinárodnej federácie konzultačných inžinierov (FIDIC).

Aby sa urýchlila projektová príprava, niekedy sa dokumentácie DSP a DRS spájajú do dokumentácie na stavebné povolenie v podrobnostiach na realizáciu stavby (DSPRS), ktorá tiež stanovuje odhad nákladov na úrovni výkazov výmer. Do prílohy A sme vložili obr. A.1 s podrobným opísaním procesu plánovania a prípravy investičných projektov na Slovensku.

Verejné obstarávanie na realizáciu stavby je možné zahájiť z akejkoľvek etapy projektovej prípravy, pričom PHZ sa stanovuje na základe odhadu nákladov najvyššej projektovej prípravy, ktorú projekt má vypracovanú. Na rozdiel od DRS a DSPRS sa pri nižších stupňoch projektovej prípravy, akými sú DSP, DÚR alebo ŠR, používa tzv. žltá kniha FIDIC. V rámci žltej knihy FIDIC zhotoviteľ stavby zodpovedá aj za vypracovanie chýbajúcich vyšších stupňov projektovej prípravy a PHZ je chápaná ako paušálna cena za realizáciu celej stavby vrátane chýbajúcej projektovej dokumentácie. Žltá kniha FIDIC môže priniesť výhody v podobe rýchlejšej výstavby, menšieho počtu dodatkov k zmluvám a možnosti pre zhotoviteľa nájsť efektívnejšie riešenia. V prípade metódy podľa žltej knihy FIDIC sú uchádzači konfrontovaní s problematikou hodnotenia rizík, ktoré sú im pripísané zo strany zadávateľa. Hodnotenie týchto rizík je potrebné zohľadniť v ponukovej cene, ktorá bude, v prípade úspechu vo verejnom obstarávaní, uhradená zadávateľom napriek skutočnosti, že niektoré z ocenených rizík sa v priebehu výstavby nemusia prejaviť.[2]

Projekty s hodnotou nad 1 mil. € s DPH podliehajú ešte pred verejným obstarávaním hodnoteniu ÚHP[32] a to bez ohľadu na to, či sa verejné obstarávanie uskutoční formou červenej knihy FIDIC alebo formou žltej knihy FIDIC. Súčasťou hodnotenia Oddelením dopravných projektov ÚHP je aj kontrola predpokladaných jednotkových cien (tzv. benchmarkovanie) železničných investičných projektov, nakoľko narábanie s verejnými financiami si vyžaduje aj nestrannú kontrolu, či sú tieto zdroje skutočne vynakladané efektívne. Doterajšia prax na Oddelení dopravných projektov ÚHP spočíva v benchmarkovaní pomocou cenových databáz *Odborový triednik stavebných konštrukcií a prác* (OTSKP) a *Zborník pre údržbu a opravy železničnej infraštruktúry* (ÚOŽI), ktoré vytvoril a aktualizuje Štátny fond dopravnej infraštruktúry (SFDI) spadajúci pod Ministerstvo dopravy Českej republiky, keďže v čase písania predloženej diplomovej práce nedisponujú s vlastnou databázou jednotkových cien pre železničné investičné projekty[23].

Je žiadané, aby bola PHZ vypracovaná s čo najrealistickejšími cenami. Pri systema-

tický vyššej PHZ sa vysielala signál trhu o neschopnosti štátu odhadnúť cenu. To vytvára priestor napr. pre potenciálne kartelové dohody medzi uchádzačmi vo verejnom obstarávaní. Pri príliš nízkej PHZ môže štát alokovať až príliš nízku sumu finančných prostriedkov pre investičný projekt a ďalšie financie môže byť problematické získať. Zle nastavená PHZ môže negatívne ovplyvniť aj výber projektu a bráni efektívnejšiemu čerpaniu eurofondov. Tie sú blokované vo výške odvodená od PHZ až do ukončenia projektu. Ak sa teda vysúťažá značne nižšia cena, ako bola PHZ, navyše blokované eurofondy sa už nemusia stihnúť vyčerpať na iné projekty a musia sa vrátiť.[31]

1.2 Rozpočty železničných investičných projektov

Výkazy výmer predpokladaných, ako aj zmluvných rozpočtov železničných investičných projektov sú rozdelené na stavebné objekty (SO) a prevádzkové súbory (PS), čo sú množiny pozostávajúce z položiek stavebných prác a materiálov. Stavebný objekt je „*priestorovo ucelená alebo aspoň technicky samostatná časť stavebnej časti stavby*“[35] (napr. výpravná budova, železničný zvršok, trakčné vedenie a i.). Prevádzkový súbor je „*súhrn strojov a zariadení vrátane ich montáží a inventára investičného charakteru, ktorý slúži na zabezpečenie čiastkového samostatného technologického procesu základnej alebo pomocnej výroby, prípadne netechnologického procesu a uvádza sa do prevádzky v súvislom čase*“[35] (napr. železničné zabezpečovacie zariadenia, oznamovacie zariadenia, kamerový systém a i.). Pre zjednodušenie sme ďalej nazývali stavebné objekty a prevádzkové súbory spoločným názvom *objekty*, ktoré sme roztriedili do tab. 1.1 na str. 18 podľa profesijných odborov.

Počet objektov v rozpočte priamoúmerne závisí od rozdelenia projektu na menšie časti (zväčša na kratšie medzistaničné traťové úseky, zastávky a stanice), pričom každá takáto časť projektu obsahuje rovnaké typy objektov, ktoré zvyknú byť v rozpočte navyše uvádzané v samostatných hárkoch v programe Microsoft Excel[23]. Kým malé projekty tak zvyknú mať počet objektov rádovo v jednotkách, veľké projekty (nad 40 mil. € s DPH) ich môžu mať až stovky práve z dôvodu rozdelenia projektu na menšie časti.

Každý objekt obsahuje množstvo položiek stavebných prác a materiálov, ktoré majú priradené unikátne kódy zväčša podľa *Triednika stavebných konštrukcií a prác* (TSKP) slúžiace na presnú identifikáciu danej položky. Prvých 5 číslíc podľa TSKP je povinných

Tab. 1.1: Roztriedenie objektov podľa profesijných odborov

Typ odboru	Kód odboru	Obsah odboru
PS	21	Železničné zabezpečovacie zariadenia
PS	22	Oznamovacie zariadenia
PS	23	Dielenská technológia
PS	24	Silnopráúdová technológia
PS	25	Radiofikácia
PS	26	Elektrická požiarňa signalizácia
PS	27	Poplachový systém narušenia a kamerový systém
PS	28	Integrovaný diagnostický systém
PS	29	Kontrola a riadenie technologického procesu napájania elektrifikovaných tratí
SO	31	Príprava územia
SO	32	Železničný zvršok a spodok, nástupištia, káblová chráničková trasa, priepusty
SO	33	Mosty a umelé stavby
SO	34	Pozemné stavby (budovy, prístrešky, protihlukové steny a i.)
SO	35	Trakčné vedenie a energetika
SO	36	Slabopráúdové rozvody
SO	37	Inžinierske siete
SO	38	Cesty a prístupové komunikácie
SO	39	Ostatné úpravy

Zdroj: [23]

s fixnými pravidlami ich určenia a obsahujú informácie o „*podrobnom členení produkcie stavebníctva*“[5]. Nedokážu však dostatočne podrobne identifikovať konkrétnu položku, preto sa rozširujú o ďalšie číslce, pre ktoré však neexistujú fixne stanovené pravidlá.

TSKP sa používa na vypracovanie noriem, ako sú ceny stavebných prác a materiálov. Na základe týchto informácií sa vytvárali a stále vytvárajú rozpočty, výrobné kalkulácie a faktúry väčšinou firmami pôsobiacimi na slovenskom stavebnom trhu[5]. Pri tvorbe rozpočtov sa používajú rôzne komerčné softvéry obsahujúce cenové databázy, ktorých číslovanie je založené práve na TSKP. V prípade železničných investičných projektov ide najčastejšie o cenové databázy CENEKON a ODIS[15]. Kvôli necentralizovanému systému TSKP sa však stáva, že jedna cenová databáza má pod určitým kódom definovanú inú položku, ako iná databáza. Taktiež nie všetky stavebné práce a materiály sa nachádzajú v jednotlivých cenových databázach, nakoľko na stavebnom trhu dochádza k neustálemu vývoju a zmenám[5].

Napriek tomu, že TSKP pokrýva široké spektrum stavebných prác a materiálov, chýbajú v ňom niektoré druhy prác, ako napr. dokončovacie práce, úpravy povrchov a výplne otvorov. Tiež nie sú zahrnuté niektoré stavebné remeselnícke práce. Z položiek relevantných pre železničné investičné projekty chýbajú práce a materiály najmä pre železničné zabezpečovacie zariadenie, rádiofikáciu tratí a i. Rôzni spracovatelia smerných orientačných jednotkových cien sa snažia tieto práce dopĺňať do svojich databáz, ale bez vzájomnej koordinácie, takže v praxi dochádza k tomu, že jedna a tá istá položka má rôzne kódy v rôznych rozpočtoch z dôvodu, že ho robila iná firma (príklad v tab. 1.2).[5, 23]

Tab. 1.2: Príklad nesynchronizácie kódov položiek

Kód položky	Názov položky	Merná jednotka
3410328801	„Kábel TCEKPFLEZE 3P 1,0“	meter
3415002002.1	„Kábel TCEKPFLEZE 3 p 1,0“	meter
22006153103	„Kábel TCEKPFLEZE 3p 1,0“	meter
2200615311	„Kábel TCEKPFLEZE 3P s jadr.Cu 1,0 mm“	meter
227006990310	„Kábel TCEKPFLEZE 3 p 1,0“	meter
2276990310	„Kábel TCEKPFLEZE 3P 1,0“	meter

Zdroj: [23]

V systéme TSKP taktiež nie sú stanovené merné jednotky jednotlivých položiek. Používané skratky merných jednotiek nie sú ani centralizované. Dochádza tak k tomu, že jedna merná jednotka je v rôznych rozpočtoch uvádzaná s inou skratkou, napr. merná jednotka s celým názvom *súbor* má skratky v rozpočtoch železničných investičných projektov ako *sub*, *súb*, *sub.*, *súb.*, *subor*, *súbor*, atď. Mnohé merné jednotky nie sú ani súčasťou Medzinárodnej sústavy jednotiek, čo môže komplikovať benchmarkovanie položiek s takýmito mernými jednotkami.[23]

Napriek svojim nedostatkom (t. j. uzavretý pre nové zmeny a nevyvážený hierarchický systém s nekompatibilným kódovaním) sa TSKP stále v Slovenskej a Českej republike používa najmä v súkromnom sektore, a to kvôli „zaužívanej tradícii“ v stavebnej praxi[5]. Aj keď Slovenská a Česká republika prevzali systém TSKP z čias spoločnej Československej socialistickej republiky, po rozpade spoločného štátu sa kódovanie výrazne medzi sebou odlišilo a nie je porovnateľné. Aj preto benchmarkovanie oceneného výkazu výmer českými

cenovými databázami má určité nedostatky. Nie je možné proces automatizovať, pretože položky v českom systéme nie sú bijektívne ku slovenskému systému TSKP a niektoré položky nie je možné vôbec benchmarkovať[23]. ÚHP tieto nedostatky rieši manuálnou kontrolou položiek podľa názvu a nie podľa číselného kódu, čo je ale časovo náročné s určitým rizikom nepresnosti, pričom časová náročnosť benchmarkovania môže ovplyvniť kvalitu hodnotenia, nakoľko má ÚHP povinnosť v zmysle § 19a zákona č. 523/2004 Z. z.[38] vypracovať ekonomické hodnotenie projektu do 30 dní.

V súčasnosti pri benchmarkovaní predpokladaného rozpočtu železničného projektu musí ÚHP riešiť okrem vyššie zmienených problémov aj prevedenie jednotkových cien benchmarku na spoločnú cenovú úroveň s ocenenými výkazmi výmer, tzv. indexácia cien. Metodika, ktorou indexujú Železnice Slovenskej republiky (ŽSR) ocenené výkazy výmer od projektantov v prípade neskoršieho predloženia projektu na hodnotenie ÚHP, než boli tieto výkazy výmer ocenené, nemá jednotné pravidlá. Preto sa proces indexácie od projektu k projektu líši v závislosti od toho, kto ho indexoval. ŽSR pri indexácii prihlasujú individuálne na každý objekt a to aj jednotlivo na časti objektu. Takže časť jedného objektu môže byť indexovaná iným indexom, ako iná časť toho istého objektu. Pre potreby našej práce sme si tento spôsob indexácie nazvali *podľa charakteru objektu*. ŽSR využívajú pre svoj spôsob indexácie podľa charakteru objektu až 6 indexov cien stavebných prác podľa klasifikácie stavieb definovaných Štatistickým úradom Slovenskej republiky (ŠÚ SR), a to[22]:

- 124 Budovy pre dopravu a telekomunikáciu,
- 211 Cestné a miestne komunikácie,
- 212 Železnice a dráhy,
- 214 Mosty, nadjazdy, tunely a podzemné dráhy,
- 221 Diaľkové potrubné, telekomunikačné a elektrické rozvody,
- 222 Miestne potrubné a káblové rozvody.

Príkladov nejednotného indexovania spôsobom podľa charakteru objektu je mnoho, napr.[22]:

- Poplachový systém narušenia má kód odboru PS 27, pričom v niektorých projektoch je indexovaný podľa indexu 212 Železnice a dráhy a v iných podľa indexu 222 Miestne potrubné a káblové rozvody.
- Káblová chráničková trasa má kód odboru SO 32, pričom v niektorých projektoch je indexovaná podľa indexu 212 Železnice a dráhy a v iných podľa indexu 222 Miestne potrubné a káblové rozvody.
- Prístrešky pre cestujúcich majú kód odboru SO 34, pričom v niektorých projektoch sú indexované podľa indexu 124 Budovy pre dopravu a telekomunikáciu a v iných podľa indexu 212 Železnice a dráhy.

ÚHP pre potreby indexovania jednotkových cien benchmarku tento spôsob indexovania z dôvodu synchronizácie so ŽSR prevzal, avšak kvôli zjednodušeniu a jednotnému používaniu naprieč projektami ho modifikoval na systém, ktorý sme pre potreby našej práce nazvali *systém kódov odborov*. Priradenie jednotlivých indexov ku kódom odborov podľa indexovania systémom kódov odborov sme spísali do tab. 1.3.

Tab. 1.3: Priradenie indexov ku profesijným odborom podľa systému kódov odborov

Typ odboru	Kód odboru	Obsah odboru	Index
PS	21	Železničné zabezpečovacie zariadenia	212 Železnice a dráhy
PS	22	Oznamovacie zariadenia	212 Železnice a dráhy
PS	23	Dielenská technológia	212 Železnice a dráhy
PS	24	Silnoprúdová technológia	212 Železnice a dráhy
PS	25	Radiofikácia	212 Železnice a dráhy
PS	26	Elektrická požiarňa signalizácia	212 Železnice a dráhy
PS	27	Poplachový systém narušenia a kamerový systém	212 Železnice a dráhy
PS	28	Integrovaný diagnostický systém	212 Železnice a dráhy
PS	29	Kontrola a riadenie technologického procesu napájania elektrifikovaných tratí	212 Železnice a dráhy
SO	31	Príprava územia	212 Železnice a dráhy
SO	32	Železničný zvršok a spodok, nástupištia, káblová chráničková trasa, priepusty	212 Železnice a dráhy
SO	33	Mosty a umelé stavby	214 Mosty, nadjazdy, tunely a podzemné dráhy
SO	34	Pozemné stavby (budovy, prístrešky, protihlukové steny a i.)	124 Budovy pre dopravu a telekomunikáciu
SO	35	Trakčné vedenie a energetika	212 Železnice a dráhy
SO	36	Slaboprúdové rozvody	222 Miestne potrubné a káblové rozvody
SO	37	Inžinierske siete	222 Miestne potrubné a káblové rozvody
SO	38	Cesty a prístupové komunikácie	211 Cestné a miestne komunikácie
SO	39	Ostatné úpravy	211 Cestné a miestne komunikácie

Zdroj: [23]

Mnohé rozpočty však nemajú v číslach objektu uvedené kódy odborov, nakoľko to nie je povinný údaj, takže sa musia priradovať manuálne, čo opäť sťažuje a spomaľuje prácu

ÚHP. Definícia viacerých profesijných odborov navyše môže spadať pod viacero typov indexov na základe klasifikácie stavieb, teda použitý index na základe uvedeného kódu odboru (v prípade, ak je uvedený) nemusí vždy zodpovedať tomu, čo daná klasifikácia stavby skutočne aj obsahuje. Napr. kým kód odboru 32 obsahuje železničný zvršok a spodok, ktoré jasne patria pod klasifikáciu stavby 212 Železnice a dráhy, obsahuje taktiež aj priepusty, ktoré by mali charakterom patriť skôr pod klasifikáciu stavby 214 Mosty, nadjazdy, tunely a podzemné dráhy[23].

Aj zmluvné rozpočty sa indexujú, a to podľa inflačnej doložky v zmluve z dôvodu nárastu cien stavebných prác a materiálov, ktorá je od roku 2022 definovaná metodickým pokynom č. 19/2022 vydaného MD SR, „*ktorým sa stanovuje mechanizmus úpravy ceny v dôsledku zmien nákladov pri projektoch opravy a údržby, výstavby, modernizácie a rekonštrukcie inžinierskych stavieb a budov*“[16]. Stanovený vzorec pre výpočet indexácie (1.1) obsahuje dve časti. Prvá, fixná časť, predstavuje 10 % z celkovej realizovanej stavby a náklady v tejto časti nepodliehajú indexácii. Druhá časť, ktorá predstavuje 90 % z celkových nákladov, podlieha indexácii. Vzorec pre výpočet indexácie podľa tohto metodického pokynu je definovaný ako[16]

$$P_t = 0,10 + 0,20 \cdot \left(\frac{HICP_t}{HICP_{t_0}} \right) + 0,08 \cdot \left(\frac{D_t}{D_{t_0}} \right) + 0,62 \cdot \left(\frac{CMI_t}{CMI_{t_0}} \right), \quad (1.1)$$

kde

- koeficient zmeny P_t sa aplikuje na odhadovanú hodnotu zmluvy za kvartál t a slúži ako faktor úpravy,
- t je ukončený kvartál (koncový), teda rozhodujúce obdobie, za ktoré zhotoviteľ uplatňuje indexáciu,
- referenčné obdobie t_0 predstavuje kvartál, v ktorom bol kalendárny deň, keď lehota na predkladanie ponúk do súťaže na zhotovenie stavby vypršala,
- koeficient 0,10 predstavuje fixnú časť výdavkov na stavebné práce a stavby, ktoré nie sú zahrnuté v indexácii,
- koeficient 0,20 zodpovedá 20 % úprave, ktorá reflektuje zmeny nákladov na prácu alebo osobné náklady v súvislosti s uskutočnenými stavebnými činnosťami a stavbami, ktoré podliehajú indexácii,

- *HICP* je ukazovateľ *Harmonizované indexy spotrebiteľských cien* v Slovenskej republike,
- koeficient 0,08 predstavuje 8 % časť nákladov na vykonané stavebné práce a stavby, ktoré podliehajú cenovej úprave, a zohľadňuje zmenu cien pohonných hmôt, konkrétne motorovej nafty,
- *D* je ukazovateľ *Priemerné ceny pohonných látok* v Slovenskej republike (motorová nafta),
- koeficient 0,62 zastupuje 62 % podiel nákladov za stavebné práce a stavby, ktoré podliehajú cenovej úprave a reflektuje zmenu nákladov na ceny materiálov a produktov spotrebovaných v stavebníctve v Slovenskej republike,
- *CMI* je ukazovateľ *Indexy cien stavebných prác a materiálov* za štvrtrok.

Systém indexácie podľa metodického pokynu č. 19/2022 však neprihliada nielen na špecifické objekty v zmysle klasifikácie stavieb, ale ani na typ projektu. Indexuje rovnakým indexom investičné projekty ako železničnej infraštruktúry, tak aj cestnej infraštruktúry či pozemných stavieb[16].

2 Potrebné poznatky zo štatistiky

Problematika odhadovania dát je v literatúre široko popísaná a jej podstatou je čo najpresnejšie sa priblížiť ku skutočným hodnotám. Spresňovanie odhadu sa môže robiť viacerými spôsobmi. Pre potreby našej práce sme v tejto kapitole popísali zaužívané matematicko-štatistické metódy na identifikáciu nestability v dátach, na identifikáciu vychýlenosti dát, na identifikáciu outlierov (vybočujúcich hodnôt) v dátach, ako aj samotné metódy odhadovania bodových hodnôt, validáciu týchto odhadov a určenie ich presnosti.

Nakoľko už popis dát v kapitole 1 naznačil, že dáta, s ktorými sme v našej práci pracovali, boli nezávislé a vybrané z jednorozmerného rozdelenia, pre metódy popísané v tejto kapitole sme preto predpokladali nasledovné: Nech $X_n = \{x_1, x_2, \dots, x_n\}$ je množina dát nezávislé vybraných z jednorozmerného rozdelenia, ktoré sú k dispozícii a nech $Y_n = \{y_1, y_2, \dots, y_n\}$ je iná množina jednorozmerných dát, ktoré majú byť odhadované.

2.1 Metódy na identifikáciu nestability v dátach

Dôležitou súčasťou analýzy dát, ako aj spresňovania odhadov bodových hodnôt je detegovanie nestability v dátach, ktorá je kľúčová pre mnohé rozhodovacie procesy. Vo všeobecnosti sa považujú dáta s vysokou mierou nestability za znakom problémov v procesoch, ktoré tieto dáta generujú. Z tohto dôvodu bolo dôležité identifikovať a analyzovať nestabilitu v dátach, aby sme pochopili jej príčiny a mohli prijať korekčné opatrenia.

Zaužívanou metódou v štatistike na identifikovanie nestability v dátach je štandardná odchýlka, ktorá sa upravuje pre vybranú časť dát ešte Besselovou korekciou na výberovú štandardnú odchýlku s definovanú ako[12]

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

kde n je počet hodnôt v množine X_n , x_i je i -tá hodnota z množiny X_n a $\bar{x}$ je aritmetický priemer hodnôt z množiny X_n .

Pre zistenie relatívneho rozptylu vybranej časti dát sa používa výberový koeficient variácie CV definovaný ako[12]

$$CV = \frac{s}{\bar{x}},$$

kde s je výberová štandardná odchýlka hodnôt z množiny X_n a $\bar{x}$ je aritmetický priemer hodnôt z množiny X_n .

2.2 Metódy na identifikáciu vychýlenosti dát

Ďalším dôležitým faktorom ovplyvňujúcim výber vhodných metód na odhadovanie dát, ako aj ovplyvňujúci ich výsledky, je vychýlenosť dát. Najistejším spôsobom zistenia vychýlenosti dát je vizualizácia dát, napr. histogramom, ktorá poskytuje rýchly prehľad o tvare dátového rozdelenia.[26]

V prípade, že vizualizácia dát nie je možná, používajú sa rôzne koeficienty indikujúce vychýlenosť dát. Zaužívaným koeficientom indikujúcim vychýlenosť dát je koeficient vychýlenosti (angl. *skewness*) S , ktorý je v štatistike známou mierou asymetrie rozdelenia pravdepodobnosti pre hodnoty náhodnej premennej a pre vybranú časť dát je definovaný ako[26]

$$S = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s^3}, \quad (2.1)$$

kde n je počet hodnôt v množine X_n , x_i je i -tá hodnota z množiny X_n , $\bar{x}$ je aritmetický priemer hodnôt z množiny X_n a s je výberová štandardná odchýlka hodnôt z množiny X_n .

Koeficient vychýlenosti sa dá interpretovať nasledovne[26]:

- Ak $S = 0$, tak dáta sú symetrické.
- Ak $S < 0$, tak dáta sú negatívne vychýlené, t. j. majú dlhý ľavý chvost.
- Ak $S > 0$, tak dáta sú pozitívne vychýlené, t. j. majú dlhý pravý chvost.

2.3 Metódy na identifikáciu outlierov v dátach

Okrem nestability a vychýlenosti môžu odhadovanie dát výrazne ovplyvniť aj outlieri. Zaužívanou metódou na identifikáciu outlierov v dátach je Tukeyho metóda, ktorá je známy grafický nástroj na zobrazenie informácií jednorozmerných dát, akými sú medián, dolný kvartil (Q1), horný kvartil (Q3), dolný extrém a horný extrém. Dolný extrém a horný extrém sa vnímajú ako hranice, ktoré určujú rozsah, v ktorom sa nachádzajú potenciálne outlieri.[11]

Tukeyho metódou sa identifikujú outlieri nasledovne[11]:

1. Najprv sa definuje vnútorný kvartilový rozsah (IQR) ako vzdialenosť medzi dolným a horným kvartilom.

2. Potom sa lokalizujú vnútorné hranice vo vzdialenosti 1,5-násobku IQR pod Q1 a nad Q3 ako $[Q1 - 1,5 \cdot IQR; Q3 + 1,5 \cdot IQR]$.
3. Následne sa lokalizujú vonkajšie hranice vo vzdialenosti 3,0-násobku IQR pod Q1 a nad Q3 ako $[Q1 - 3,0 \cdot IQR; Q3 + 3,0 \cdot IQR]$.
4. Nakoniec hodnotu medzi vnútornými a vonkajšími hranicami možno považovať za potencionálneho outliera. Hodnota prevyšujúca vonkajšie hranice je extrémny outlier.

Tukeyho metóda nevyžaduje žiadne predpoklady o distribúcii a nezávisí na priemere alebo štandardnej odchýlke. Avšak Tukeyho metóda nemusí byť vhodná pre malú vzorku dát[25].

Hoci Tukeyho metóda môže byť vhodná pre symetrické dáta, čím viac sú dáta vychýlené, tým viac meraní môže byť identifikovaných ako outlier[34]. Je to dôsledkom založenia metódy na robustných meraniach, akými sú dolný a horný kvartil a IQR, bez ohľadu na vychýlenosť dát[11]. Preto Vanderviere a Huber v [34] predstavili metódu Adjusted Boxplot, ktorá berie do úvahy aj robustnú mieru vychýlenosti pre vychýlenú distribúciu, tzv. mediánový pár hodnôt (angl. *medcouple*) (MC). MC dát je definovaný ako

$$MC(x_1, x_2, \dots, x_n) = med \frac{(x_j - med_k) - (med_k - x_i)}{x_j - x_i},$$

kde med_k je definovaný ako medián hodnôt z množiny X_n , i a j spĺňajú podmienku $x_i \leq med_k \leq x_j$ a $x_i \neq x_j$, pričom platí, že $x_1 \leq x_2 \leq \dots \leq x_n$.

Potom interval pre vnútorné a vonkajšie hranice metódy Adjusted Boxplot je definovaný ako

$$[L, U] = \begin{cases} [Q_1 - 1,5 \cdot \exp(-3,5MC) \cdot IQR; Q_3 + 1,5 \cdot \exp(4,0MC) \cdot IQR], & \text{ak } MC \geq 0, \\ [Q_1 - 1,5 \cdot \exp(-4,0MC) \cdot IQR; Q_3 + 1,5 \cdot \exp(3,5MC) \cdot IQR], & \text{ak } MC \leq 0, \end{cases}$$

kde L je vnútorná hranica a U je vonkajšia hranica intervalu. Pozorovania mimo tohto intervalu sú považované za outliere.

Interval možných hodnôt MC je z $[-1, 1]$ a jeho interpretácia je nasledovná[34]:

- Ak $MC = 0$, tak dáta sú symetrické a metóda Adjusted Boxplot = Tukeyho metóda.
- Ak $MC > 0$, tak dáta majú vpravo vychýlenú distribúciu.
- Ak $MC < 0$, tak dáta majú naľavo vychýlenú distribúciu.

2.4 Metódy na odhadovanie bodovej hodnoty

Existuje mnoho samotných metód na odhadovanie bodovej hodnoty. V tejto kapitole sme popísali tie bežne zaužívané v štatistike, pričom sme vychádzali voľne najmä zo zdrojov [18, 21].

Základnou štatistickou metódou na odhadnutie skutočnej hodnoty zo súboru dát je aritmetický priemer $\bar{x}$, ktorý je pre vybranú časť dát definovaný ako

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

kde n je počet hodnôt v množine X_n a x_i je i -tá hodnota z množiny X_n .

Avšak aritmetický priemer môže byť výrazne ovplyvnený outliermi v súbore dát, preto nemusí byť najvhodnejší pre isté typy dát.

V prípadoch, keď nie sú všetky hodnoty v súbore dát rovnako dôležité, sa aritmetický priemer modifikuje na vážený priemer $\bar{x}_w$, ktorý je pre vybranú časť dát definovaný ako

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}, \text{ pričom } \sum_{i=1}^n w_i > 0,$$

kde n je počet hodnôt v množine X_n , x_i je i -tá hodnota z množiny X_n a w_i je i -tá váha priradená k i -tej hodnote x_i . Týmto spôsobom sa uprednostňujú niektoré hodnoty na úkor iných, čím sa môže dosiahnuť správnejší odhad reálnej hodnoty zo súboru dát v prípadoch, keď nie sú všetky hodnoty v súbore dát rovnako dôležité, pretože hodnotám s väčšou váhou sa priraduje aj väčší vplyv na odhad skutočnej hodnoty.

Ďalšou používanou metódou na odhadnutie skutočnej hodnoty zo súboru dát je medián. Medián je hodnota v usporiadanom súbore dát, ktorá sa nachádza v strede súboru, takže polovica dát je nad ňou a polovica pod ňou a pre vybranú časť dát je definovaný v závislosti od počtu hodnôt v množine X_n ako

$$med = x_{((n+1)/2)},$$

ak n je nepárny počet hodnôt v množine X_n a

$$med = \frac{x_{(n/2)} + x_{((n/2)+1)}}{2},$$

ak n je párny počet hodnôt v množine X_n , kde n je počet hodnôt v množine X_n a $x_{(1)} \leq \dots \leq x_{(n)}$ sú zoradené hodnoty z množiny X_n .

Keďže medián je robustná metóda, t. j. nie je citlivá na outliere v dátach, používa sa v prípadoch, keď by outliere mohli ovplyvniť výsledok odhadu. Výhodou mediánu je, že ak sa v dátach nachádzajú outliere, ktoré by mohli ovplyvniť výsledok odhadu, tak medián týmto hodnotám neprisudzuje veľkú váhu a výsledok odhadu bude bližší k skutočnej hodnote.

Alternatívou k priemeru a mediánu sa ešte zvykne v štatistike používať modus, ktorý opisuje najčastejšie sa vyskytujúcu hodnotu v súbore dát. Inými slovami, je to hodnota s najväčšou frekvenciou. Modus je však nepoužiteľný pri odhadovaní skutočnej hodnoty zo súboru dát v prípade, ak je v súbore dát viacero hodnôt s rovnakou najväčšou frekvenciou. Taktiež pri vychýlenosti dát sa môže modus posunúť od zvyčajnej polohy, ktorú zaujíma v nevychýlenom súbore. V prípade, že sú v súbore dát vychýlené hodnoty, ktoré sa vyskytujú výrazne častejšie ako zvyšné hodnoty, modus sa môže posunúť smerom k týmto vychýleným hodnotám. Naopak, ak sú v súbore dát vychýlené hodnoty, ktoré sa vyskytujú veľmi zriedka, modus sa môže posunúť smerom k zvyšným hodnotám v súbore.

2.5 Metódy na validovanie odhadu modelom

Výber a nastavenie vhodného modelu na odhadovanie bodovej hodnoty si vyžaduje validovanie odhadu týmito modelmi. Štandardným nástrojom pre validovanie odhadu nejakým modelom je rozdelenie databázy na dve dátové sady – trénovaciu a testovaciu. V prípade, že je potrebné aj nastaviť hyperparametre modelu, databáza sa zvykne rozdeliť až na tri dátové sady, a to trénovaciu, validačnú a testovaciu.[29]

Trénovacia dátová sada predstavuje podmnožinu pôvodnej dátovej sady, ktorá slúži na trénovanie modelu. Model sa učí na základe vlastností a vzťahov v trénovacej dátovej sade.[29]

Validačná dátová sada je ďalšia podmnožina pôvodnej dátovej sady, ktorá slúži na overenie a nastavenie modelu počas jeho trénovania. Na validačnej dátovej sade sa testujú rôzne hyperparametre a modelové architektúry, aby sa zlepšila presnosť a generalizovateľnosť modelu.[29]

Testovacia dátová sada je tretia podmnožina pôvodnej dátovej sady, ktorá slúži na overenie presnosti modelu po jeho trénovaní a nastavení na trénovacej a validačnej dátovej

sade. Testovacia dátová sada by mala byť nezávislá na trénovacej a validačnej dátovej sade a mala by simulovať reálne podmienky, na ktoré bude model aplikovaný.[29]

Cieľom rozdelenia databázy na trénovaciu, validačnú a testovaciu dátovú sadu je minimalizovať riziko preučenia (angl. *overfitting*) modelu na trénovacej dátovej sade a maximalizovať jeho generalizovateľnosť na nové dáta. Tento proces pomáha aj odhaliť problémy v dátach a v modeli ešte pred jeho nasadením do praxe.[29]

Rozdeliť databázy na trénovaciu a testovaciu, prípadne trénovaciu, validačnú a testovaciu dátovú sadu je možné viacerými metódami, a to buď spôsobom obvyčajnej validácie, alebo krížovej validácie. Obyčajná validácia rozdelí databázu v určitom pomere, buď na základe vopred stanoveného kritéria, alebo na základe subjektívneho úsudku. Výhodou tohto spôsobu je napr. zmenšenie časovo výpočtovej náročnosti. Tento spôsob má však niekoľko nevýhod najmä pre malé množstvo dát, pretože model môže byť ovplyvnený náhodnými vlastnosťami testovacej dátovej sady, čo môže viesť k nedostatočnej validácii. Taktiež dochádza k istej miery skreslenosti výsledkov, keďže dáta boli vyskúšané iba na jednej množine dát.[7]

Krížová validácia na rozdiel od obvyčajnej validácie používa viacero testovacích a trénovacích dátových sád, čím sa zabezpečí robustnejšia a spoľahlivejšia validácia výkonnosti modelu a znižuje sa riziko ovplyvnenia výsledkov náhodnými vlastnosťami testovacích dát a zvyšuje sa spoľahlivosť výsledkov.[7]

Existuje niekoľko typov krížovej validácie, z ktorých najpoužívannejšie sú[7]:

- k -násobná krížová validácia (angl. *k-fold cross-validation*) – Dátový súbor sa rozdelí do k disjunktných množín rovnakej veľkosti. k -krát sa vykoná trénovanie modelu na zvyšných $k - 1$ množinách a testovanie na zostávajúcej množine.
- Odlož-jeden-mimo (angl. *leave-one-out*) krížová validácia – Pozorovania sa po jednom permutačne postupne použijú ako testovacia množina a zostávajúce pozorovania ako trénovacia množina.
- Náhodné rozdelenie (angl. *random subsampling*) – Dátový súbor sa náhodne rozdelí na trénovaciu a testovaciu množinu. Postup sa opakuje pre viacero náhodných rozdelení.

2.6 Metódy na meranie presnosti odhadu modelom

Pre umožnenie porovnať a vybrať vhodný model na odhadovanie bodovej hodnoty je dôležité určiť presnosť odhadu modelom. Na meranie presnosti odhadu nejakým modelom existuje v štatistike viacero metrík. V tejto kapitole sme vychádzali voľne zo zdrojov [9, 10].

Medzi najzaužívannejšie metriky na meranie presnosti patrí druhá odmocnina strednej kvadratickej odchýlky (angl. *root mean squared error*) (RMSE) definovaná ako

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2\right)},$$

kde n je počet hodnôt v množine X_n , y_i je i -tá hodnota z množiny Y_n a $\hat{y}_i$ je odhadovaná hodnota i -tej hodnoty y_i .

Výhodou RMSE je ľahká interpretácia z dôvodu zachovania merných jednotiek, t. j. vracia chybu v merných jednotkách, v ktorých sa aj premenná pohybuje. Alternatívnou metrikou na meranie presnosti v absolútnych chybách je stredná absolútna odchýlka (angl. *mean absolute error*) (MAE) definovaná ako

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

kde n je počet hodnôt v množine X_n , y_i je i -tá hodnota z množiny Y_n a $\hat{y}_i$ je odhadovaná hodnota i -tej hodnoty y_i .

MAE sa používa v situáciach, keď outliere by mohli ovplyvniť výsledok a výpočet RMSE by bol veľmi skreslený, pretože sa nezohľadňuje jej druhá mocnina. Rovnako ako RMSE sa výsledky MAE dajú ľahko interpretovať, pretože merné jednotky sú rovnaké, v akých sa aj premenná pohybuje.

Kým RMSE a MAE sú metriky založené na meraní chýb v absolútnych mierkach, existujú aj metriky založené na meraní chýb v relatívnych mierkach, napr. stredná absolútna percentuálna odchýlka (angl. *mean absolute percentage error*) (MAPE), ktorá je definovaná ako

$$MAPE = \frac{100 \%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i},$$

kde n je počet hodnôt v množine X_n , y_i je i -tá hodnota z množiny Y_n a $\hat{y}_i$ je odhadovaná hodnota i -tej hodnoty y_i .

Výhodou MAPE je, že exprimuje výchylky v percentách, čo je intuitívne zrozumiteľnejšie ako v absolútnych hodnotách a umožňuje porovnávanie presnosti odhadov medzi

rôznymi dátovými sadami bez ohľadu na ich mierku. Avšak MAPE má aj svoje nevýhody, ktoré sa prejavujú najmä v prípade, keď skutočná hodnota je blízko alebo rovná nule, čo môže spôsobiť, že percentuálna odchýlka sa stane nevypovedajúcou alebo nekonečnou.

V prípadoch, keď nie sú všetky odhadované hodnoty v súbore dát rovnako dôležité, sa MAPE modifikuje na váženú strednú absolútnu percentuálnu odchýlku (angl. *weighted mean absolute percentage error*) (wMAPE), ktorá váži absolútnu percentuálnu chybu každej predikcie podľa jej relatívnej dôležitosti alebo veľkosti a je definovaná ako

$$wMAPE = \frac{\sum_{i=1}^n w_i \cdot |y_i - \hat{y}_i|}{\sum_{i=1}^n w_i \cdot y_i} \cdot 100 \%,$$

kde n je počet hodnôt v množine X_n , y_i je i -tá hodnota z množiny Y_n , w_i je i -tá váha priradená k i -tej hodnote y_i a $\hat{y}_i$ je odhadovaná hodnota i -tej hodnoty y_i .

3 Metodika tvorby cenovej databázy a analýzy dát

Problematika odhadu nákladov železničných investičných projektov popísaná v kapitole 1 nás motivovala navrhnúť pre potreby Oddelenia dopravných projektov ÚHP **vlastný model odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov**, ktorý by bol efektívnejší než súčasný spôsob odhadovania týchto jednotkových cien Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI s vyššou alebo aspoň porovnateľnou úrovňou presnosti. Túto úlohu sme považovali za primárny cieľ našej práce. Na dosiahnutie primárneho cieľa sme potrebovali vyriešiť viaceré problémy, pre ktoré sme si vytýčili parciálne ciele a ktoré sme dosiahli zrealizovaním nasledujúcich krokov:

1. Pre **potrebu relevantných vstupných dát do modelu** sme vytvorili databázu jednotkových cien stavebných prác a materiálov zo zmluvných rozpočtov už realizovaných železničných investičných projektov, nakoľko Oddelenie dopravných projektov ÚHP nedisponovalo v čase písania našej práce žiadnou vlastnou databázou takýchto jednotkových cien.
2. Pre **rovnocennosť jednotlivých projektov vstupujúcich do modelu** sme zostrojili vlastný mechanizmus šandardizácie váh projektov, nakoľko každý projekt, veľký či malý, poskytuje hodnotovo rovnocennú informáciu o jednotkových cenách na trhu.
3. Pre **zjednotenie cenovej úrovne projektov vstupujúcich do modelu** sme navrhli vlastný systém indexácie, ktorý by bol nezávislý od nepovinných údajov v rozpočtoch a zároveň by reflektoval systém klasifikácie stavieb ŠÚ SR, podľa ktorého sa stanovuje aj samotná hodnota použitých indexov, nakoľko existujúci systém indexácie Oddelením dopravných projektov ÚHP je závislý na kódach odborov, ktoré nie vždy sú v rozpočtoch aj uvádzané a plne nereflektujú systém klasifikácie stavieb ŠÚ SR.
4. Pre **dosiahnutie skutočne relevantnej výpovednej hodnoty modelu** sme identifikovali položky s následným očistením dát od nich, u ktorých by odhadovanie jednotkových cien nebolo relevantné, nakoľko buď ich jednotková cena mala

príliš vysokú mieru fluktuácie na to, aby odhad ceny mal nejakú výpovednú hodnotu, alebo nemuseli reprezentovať to isté naprieč rôznymi projektmi, a teda ich vzájomné porovnávanie by nebolo korektné.

5. Pre **spresnenie odhadu jednotkových cien modelom** sme identifikovali outliere s následným očistením dát od nich, nakoľko by tieto vychýlené individuálne jednotkové ceny položiek mohli skresľovať výsledné odhady jednotkových cien.
6. Pre **optimalizovanie odhadovaných jednotkových cien modelom** sme matematicko-štatistickými metódami vylepšovali odhadovanie jednotkových cien, nakoľko sme chceli mať čo najpresnejší model odhadovania jednotkových cien.
7. Pre **validáciu navrhnutého modelu** sme ho aplikovali na dáta simulujúce reálne podmienky v praxi a to aj v konfrontácii s ocenenými výkazmi výmer železničných investičných projektov od projektantov, súčasným spôsobom benchmarkovania jednotkových cien stavebných prác a materiálov železničných investičných projektov Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI, ako aj s jednotkovými cenami stavebných prác a materiálov najpoužívanejšou komerčnou cenovou databázou na slovenskom stavebnom trhu CENEKON, nakoľko iba efektívnejší a aspoň porovnateľne presný model, ako tieto spôsoby odhadovania jednotkových cien, by mal aj reálny potenciál byť skutočne využívaný v praxi.

Samotnú analýzu dát sme robili v programe R a všetky nami vytvorené skripty pre potreby našej práce sme zverejnili na internetovom hostingu GitHub [27].

3.1 Tvorba cenovej databázy

Nakoľko v čase písania našej práce nedisponovalo Oddelenie dopravných projektov ÚHP žiadnou vlastnou databázou jednotkových cien železničných investičných projektov, museli sme si pre potreby našej práce takúto cenovú databázu najprv vytvoriť. Vychádzali sme zo zmluvných rozpočtov už realizovaných železničných investičných projektov zameraných na železničnú infraštruktúru, keďže len tento typ železničných investičných projektov podlieha hodnoteniu Oddelenia dopravných projektov ÚHP. Projekty zamerané na staničné budovy, informačné systémy či koľajová doprava sú charakterom výrazne odlišné, a aj preto sú hodnotené zväčša osobitne inými oddeleniami ÚHP.

Taktiež sme sa zamerali len na rozpočty projektov, ktoré boli obstarávané v stupni projektovej prípravy DRS alebo DSPRS, pretože len v tejto etape projektovej prípravy je k dispozícii podrobný rozpočet na úrovni výkazov výmer. Dôvodom tohto vymedzenia je, že nižšie etapy projektovej prípravy DSP, DÚR a ŠU, resp. projekty realizované formou žltej knihy FIDIC, neposkytujú dostatočný detail rozpočtu pre umožnenie automatizovaného benchmarkovania.

Časť zmluvných rozpočtov sme získali priamo od ŽSR[15] alebo ÚHP[23], zvyšné rozpočty boli verejne dostupné na internetových stránkach Centrálného registra zmlúv[3] a Úradu pre verejné obstarávanie[33]. Nakoľko mali rozpočty rôzne formáty a úpravu, vytvorili sme vlastný skript v programe R[27], ktorý dokázal vytiahnuť z každého rozpočtu a hárka vo formáte .xlsx iba informácie o položkách, ktoré boli pre potreby našej práce relevantné a to:

- **Názov projektu** – Názov projektu, v ktorom sa vyskytla daná stavebná položka.
- **Kód časti stavby** – Kód stavebného objektu, v ktorom sa vyskytla daná stavebná položka.
- **Názov časti stavby** – Názov stavebného objektu, v ktorom sa vyskytla daná stavebná položka.
- **Kód položky** – Kód stavebnej položky.
- **Názov položky** – Názov stavebnej položky.
- **Merná jednotka** – Skratka mernej jednotky, v ktorej bola vymieraná daná stavebná položka.
- **Množstvo** – Množstvo stavebnej položky, ktoré sa zakúpilo na stavebnom objekte.
- **Jednotková cena bez DPH** – Cena za jednu mernú jednotku danej stavebnej položky bez DPH.
- **Celková cena bez DPH** – Cena za celkové množstvo danej stavebnej položky bez DPH.

V čase odovzdávania predloženej diplomovej práce bolo v cenovej databáze spracovaných 67 železničných investičných projektov, ktorých zmluva bola, alebo sa plánovala uzavrieť v rokoch 2012 až 2023, s 281 277 stavebnými položkami v nenaindexovanej hodnote 1,62 mld. € bez DPH. Zoznam železničných investičných projektov spracovaných do našej cenovej databázy sme uviedli v prílohách B a C.

Pre prípady, ak by niektoré stavebné položky boli zaznamenané v rôznych merných jednotkách, čo by mohlo ovplyvniť zásadným spôsobom ich jednotkové ceny i napriek tomu, že by vyjadrovali tú istú stavebnú položku, sme pre zvýšenie presnosti vytvorili v našej cenovej databáze osobitný stĺpec spojením kódov položiek so skratkou merných jednotiek. Nami vytvoreným kódom sme tak identifikovali položky, u ktorých sme mali istotu, že jednotková cena zodpovedá aj danej mernej jednotke a mohli sme tak porovnávať porovnateľné.

Bez úprav by sa však v našej cenovej databáze vyskytovalo až 138 unikátnych skratiek merných jednotiek. Dôvodom bolo necentralizované používanie skratiek merných jednotiek rôznymi firmami, na čo sme upozornili aj v kapitole 1.2. Keďže sme na identifikáciu položiek používali skladbu číselného kódu a skratky mernej jednotky, tak by to mohlo znížiť presnosť odhadovania našim navrhnutým modelom z dôvodu nižšieho počtu dát pre jednotlivé položky. Tento problém sme vyriešili identifikáciou skratiek, ktoré odkazovali na tú istú mernú jednotku a následným nahradením jednotným značením. Výsledkom bolo zníženie počtu skratiek merných jednotiek z celkového počtu 138 na 55. ŽSR sme zaslali taktiež zoznam nami odporúčaných skratiek merných jednotiek, aby v budúcnosti požadovali od projektantov zapracovanie týchto skratiek merných jednotiek v rámci nimi vypracovaných rozpočtov.

Zo spoločnej databázy sme pre potreby ďalšej práce oddelili projekty nad 1 mil. € bez DPH, ktorých zmluva bola uzavretá, alebo sa plánovala uzavrieť v rokoch 2022 a 2023, ako testovaciu dátovú sadu, nakoľko tieto rozpočty boli hodnotené aj ÚHP a mali sme tak k nim k dispozícii aj ocenené výkazy výmer od projektantov, ako aj benchmark cenovými databázami SFDI, ktorý pre potreby našej práce ÚHP aktualizovalo o najnovšiu verziu cenových databáz SFDI[23].

Testovacia databáza obsahujúca len testovaciu dátovú sadu obsahovala celkovo 9 projektov s 67 494 položkami v nenaindexovanej hodnote 380,81 mil. € bez DPH, pričom bola

dostatočne diverzifikovaná vo všetkých ohľadoch. Obsahovala ako malé, tak aj veľké projekty. Projekty zamerané na železničnú trať, ako aj koľajisko v staniciach. Projekty komplexnej modernizácie, ako aj obvyčajnej obnovy. Stavebné projekty, ako aj technologické projekty a to vrátane špecifických projektov, akými sú elektrifikácia trate, implementácia globálneho systému mobilnej komunikácie pre železnice (GSM-R) či rekonštrukcia železničného tunela. Takúto diverzifikovanú dátovú sadu reprezentujúcu skutočné hodnotené železničné investičné projekty Oddelením dopravných projektov ÚHP v relatívne nedávnom období sme považovali za adekvátnu pre potreby validácie navrhnutého modelu.

Zvyšných 58 projektov sme použili ako trénovaciu databázu. Na testovacej dátovej sade sme validovali model vybraný na základe výsledkov analýz v trénovacej databáze. Zoznam projektov v trénovacej databáze sme priložili do prílohy B a zoznam projektov v testovacej databáze sme priložili do prílohy C.

3.2 Čistenie dát v cenovej databáze

Ešte pred samotnou analýzou našich dát v databáze a práce s nimi, sme museli naše dáta očistiť od nerelevantných položiek, nakoľko by s nimi závery našej analýzy a odhady jednotkových cien neboli zmysluplné. Preto sme v tejto kapitole popísali východiskové kritéria, na základe ktorých sme identifikovali nerelevantné položky v našej trénovacej databáze. Z ďalšej analýzy na trénovacej databáze sme tak vylúčili:

1. Položky s nulovou jednotkovou cenou – Z prvých pozorovaní sme si všimli, že istá malá časť položiek mala nulovú jednotkovú cenu. V zmluvách je to zdôvodnené tým, že došlo k zmene rozsahu prác medzi oceneným výkazom výmer projektantom a súťažnými podmienkami, napr. z dôvodu zmeny typu použitej technológie či materiálu, alebo danú položku zabezpečia samotné ŽSR. Dané položky sa však z rozpočtu neodstraňujú, iba sa im priradí nulová jednotková cena kvôli spätnej kompatibilite s ocenením výkazom výmer. Nakoľko táto nulová jednotková cena nereprezentuje reálnu jednotkovú cenu na trhu, skresľovali by naše výsledky.
2. Poplatky za uloženie odpadov na skládku – Sú špecifickou skupinou položiek, nakoľko do ich jednotkovej ceny vstupuje fixný poplatok daný nariadením vlády Slovenskej republiky č. 330/2018 Z. z.[20] meniaci sa v čase. Indexácia tak nemôže

zachytiť plnohodnotne ich zmenu a bolo by treba nielen v rámci našej práce, ale aj v praxi vždy individuálne prepočítat tieto poplatky.

3. Agregované položky – Tento typ položiek aj keď má rovnaký kód a mernú jednotku naprieč projektmi, vďaka agregácii nemusia popisovať vždy to isté a následne to ovplyvňuje aj ich jednotkovú cenu. Ide prevažne o položky, ktoré sú merané v merskej jednotke %, lebo tie sa odvodzujú zväčša od celkovej sumy objektu. Vynechali sme aj položky, ktoré boli uvádzané v merných jednotkách *súbor*, *komplet*, *sada* a *úsek* na základe internej analýzy ÚHP[31], ktorá zistila, že pri týchto položkách je odhadovanie optimálnej ceny náročné až nemožné z dôvodu taktiež agregovania viacerých stavebných prác a materiálov do jednej položky, ktorá naprieč projektmi nemusí vždy reprezentovať to isté.
4. Položky s kódom nezapadajúcim do TSKP – Nakoľko TSKP má povinných minimálne 5 číslic, ktorými začína každý kód položky podľa neho (ako sme to popísali v kapitole 1.2), položky, ktoré majú nižší počet číslic vo svojom kóde, alebo začínajú kód iným znakom ako číslicou, nepatria do TSKP. Takéto položky je obtiažne benchmarkovať, nakoľko v rôznych projektoch reprezentujú niečo iné, napr. kód *Pol123* doslova znamená „položka v poradí 123.“[23].

3.3 Štandardizácia váh projektov

Nakoľko každý projekt, veľký či malý, poskytuje hodnotovo rovnocennú informáciu o cenách na trhu, bolo dôležité, aby projekty mali rovnakú váhu v našej analýze a pri odhadovaní jednotkových cien, t. j. aby jeden veľký projekt, kde je jedna a tá istá položka duplikovaná viackrát na rôznych objektoch zväčša s tou istou jednotkovou cenou, neprevážila všetky ostatné menšie projekty. Z tohto dôvodu sme zostrojili vlastný mechanizmus štandardizácie váhy projektov nasledovne:

1. Najprv pomocou Tukeyho metódy s konštantou 3,0 sme identifikovali iba extrémne outliere spomedzi položiek v konkrétnom projekte, aby sme vyhodili položky, ktoré mali výrazne odlišnú jednotkovú cenu od ostatných totožných položiek v danom projekte, inak by tieto položky mohli znehodnocovať odhadovanie jednotkových cien.

2. Následne sme si previedli jednotkové ceny vo všetkých projektoch na spoločnú cenovú úroveň, aby sme ich mohli porovnávať medzi sebou.
3. Vybrali sme si všetky unikátne jednotkové ceny v rámci totožných položiek a projektov. Tieto hodnoty sme následne použili v ďalšej práci, napr. pri analýze dát a pri odhadovaní jednotkových cien.

3.4 Zjednotenie cenovej úrovne

Aby sme mohli porovnávať jednotkové ceny medzi jednotlivými zmluvnými rozpočtami pochádzajúcich z rôznych časových období, bolo potrebné ich preindexovať na spoločnú cenovú úroveň. V kapitole 1.2 sme popísali viaceré systémy indexácie rozpočtov v čase písania našej práce využívané na ŽSR, ÚHP či MD SR. Rozhodli sme sa analyzovať tie, ktoré majú jasne formulované pravidlá, t. j. systém kódov odborov a systém podľa metodického pokynu č. 19/2022.

V prípade systému kódov odborov sme pri niektorých projektoch museli kód odboru doplniť sami podľa vlastného úsudku na základe charakteristiky objektu. V prípade systému podľa metodického pokynu č. 19/2022, aby sme daný vzorec (1.1) pre výpočet indexácie mohli použiť pre naše dáta, sme si vypočítali potrebné indexy do roku 2012 a nastavili sme si, aby koeficient zmeny násobil jednotkové ceny položiek v rozpočtoch na základe času t_0 .

Nakoľko používaný systém indexácie Oddelením dopravných projektov ÚHP bol závislý na kódoch odborov, ktoré nie vždy sú v rozpočtoch aj uvádzané a plne nereflektujú klasifikáciu stavieb a metodický pokyn č. 19/2022 neprihliada špecificky nielen na klasifikáciu stavieb, ale ani len na železničné investičné projekty ako také, navrhli sme pre potreby našej práce aj vlastný systém indexácie, ktorý sme pre potreby našej práce nazvali *systém názvov objektov*. Systém názvov objektov vychádzal z kľúčových slov v názvoch objektov, čím bol nezávislý od nepovinných údajov v zmluvných rozpočtoch a zároveň mohol reflektovať aj klasifikáciu stavieb ŠÚ SR. Na samotné indexovanie sme následne použili indexy cien stavebných prác podľa klasifikácie stavieb, ktoré sa aktualizujú na štvrtročnej báze[28].

Indexy cien stavebných prác podľa klasifikácie stavieb, ktoré sme použili pre systém názvov objektov boli:

- 124 Budovy pre dopravu a telekomunikáciu,
- 212 Železnice a dráhy,
- 214 Mosty, nadjazdy, tunely a podzemné dráhy.

Vypísali sme si základy kľúčových slov, ktoré sme vytvorili na základe nášho odhadu s konzultáciou ÚHP[23], ku ktorým sme následne priradili príslušné trojčíselné kódy indexov vychádzajúc z popisu klasifikácie stavby spomínaných indexov. Nami napísaný algoritmus v programe Microsoft Excel následne prešiel všetky názvy objektov v databáze, pričom identifikoval nami stanovené kľúčové slová v názvoch objektoch. Ak kľúčové slovo nenašiel, tak priradil index 212 Železnice a dráhy. Do algoritmu sme pridali viacero podmienok, ako napr. nech pred kľúčovým slovom musela byť medzera alebo špeciálny znak a nech kľúčové slová boli vždy len základy slov, aby sme predišli neidentifikácii v prípade skloňovania daného slova. Zoznam kľúčových slov pre systém názvov objektov k indexom 124 Budovy pre dopravu a telekomunikáciu a 214 Mosty, nadjazdy, tunely a podzemné dráhy sme priložili do tab. 3.1.

Tab. 3.1: Zoznam kľúčových slov pre systém názvov objektov a im prináležiace indexy

Názov indexu	Zoznam kľúčových slov
124 Budovy pre dopravu a telekomunikáciu	budov, VB, priestor, dom, prevádzkový objekt, technologický objekt
214 Mosty, nadjazdy, tunely a podzemné dráhy	most, ŽM, estakád, viadukt, lávk, priepust, nadjazd, podjazzd, nadchod, podchod, tunel

Zdroj: vlastné spracovanie

3.5 Identifikácia nestabilných položiek v dátach

I napriek očisteniu dát od položiek popísaných v kapitole 3.2 by nie všetky položky poskytovali hodnoverný bodový odhad, nakoľko ich jednotková cena mala príliš vysokú mieru fluktuácie. Z tohto dôvodu sme potrebovali v našej práci identifikovať a následne očistiť od nich naše dáta aj položky s vysokou mierou fluktuácie, ktoré sme pre potreby našej práce nazvali *nestabilné položky*.

Spôsob, akým sme detegovali tieto nestabilné položky, bola kombinácia metód výberovej štandardnej odchýlky a výberového variančného koeficientu, ktoré sme popísali v kapitole 2.1. Kým pri položkách s nízkymi jednotkovými cenami dávalo zmysel použiť

výberový variančný koeficient na ich identifikovanie, pri položkách s vysokými jednotkovými cenami, to nebola úplne najvhodnejšia metóda a bolo lepšie použiť výberovú štandardnú odchýlku. Inými slovami, kým by bolo z pohľadu praxe akceptované, ak by sa jednotková cena pohybovala medzi 10 až 20 centami, tak by nebolo akceptované, ak by sa jednotková cena pohybovala medzi 1 mil. až 2 mil. €.

Určenie hraníc pre výberový variančný koeficient a výberovú štandardnú odchýlku, na základe ktorých sme filtrovali nestabilné položky, nebola triviálna úloha. Pri verejnom obstarávaní by mohli v praxi teoreticky nastať 3 situácie:

- Bol by len 1 uchádzač a to ten, ktorý by ponúkol najnižšiu cenu na trhu.
- Bol by len 1 uchádzač a to ten, ktorý by ponúkol najvyššiu cenu na trhu.
- Ak by bolo viac uchádzačov, výsledná cena by sa nachádzala medzi extrémami prvých dvoch situácií.

Rozdiel medzi prvými dvomi situáciami sme preto považovali za najakceptovateľnejšiu nestabilitu pre našu prácu. Interná analýza ÚHP[24] identifikovala, že rozdiel medzi najnižšou a najvyššou ponúknutou cenou za projekty ako celok vo verejnom obstarávaní je zvyčajne do 30 % za roky 2018 až 2022. Vyskytli sa však aj 2 projekty, kde tento rozdiel bol až nad 30 %, a to rovno nad 40 %. Išlo o projekty „ŽSR, Modernizácia koridoru, štátna hranica ČR/SR - Čadca - Krásno nad Kysucou (mimo), železničná trať, 3.etapa“, kde to bolo 43 % a „Rekonštrukcia Bratislavského tunela č. 2, ŽST Bratislava hl. st.“, kde to bolo 41 %. Položky s vplyvom na rozpočet nad 1 ‰ v najnižšej aj najvyššej ponuke v týchto dvoch verejných obstarávaniach sa navzájom líšili s jednotkovou cenou nanajvýš do 12-násobku. Ak by sme predpokladali, že 12-násobok je najvyšší možný akceptovateľný rozdiel, potom by všetky ostatné ponuky mali spadnúť niekde medzi ne a niekde uprostred by sme mohli očakávať takú štandardnú cenu na trhu, resp. niekedy by bola vysúťažená cena pod, niekedy nad touto strednou cenou. Ak by sme si teda vzali stred ako štandardnú cenu, potom najakceptovateľnejší výberový variančný koeficient by bol $\sqrt{12}$, keďže platí:

$$\min \cdot \sqrt{12} = \text{stred}, \text{stred} \cdot \sqrt{12} = \max \implies \max = \min \cdot \sqrt{12} \cdot \sqrt{12} = \min \cdot 12.$$

Teda položky s výberovým variančným koeficientom vyšším ako $\sqrt{12}$ by sa nemali vyskytovať štandardne na trhu, preto sme sa rozhodli ich nebenchmarkovať.

Pre výberovú štandardnú odchýlku sme aplikovali podobný postup. Položky s vplyvom na rozpočet nad 1 ‰ v najnižšej aj najvyššej ponuke vo vyššie zmienených dvoch verejných obstarávaní sa navzájom líšili s jednotkovou cenou maximálne do 50 tis. €. Ak by sme si teda vzali stred ako štandardnú cenu, potom najakceptovateľnejšia výberová štandardná odchýlka by bola 25 000, keďže platí:

$$\min + 25000 = \text{stred}, \text{stred} + 25000 = \max \implies \max = \min + 25000 + 25000 = \min + 50000.$$

Teda položky s výberovou štandardnou odchýlkou vyššiu ako 25 000 by sa nemali vyskytovať štandardne na trhu, preto sme sa rozhodli ich nebenchmarkovať.

3.6 Identifikácia outlierov v dátach

Kvôli decentralizácii kódov v TSKP spomenuté v kapitole 1.2 sme sa rozhodli preskúmať aj, či v dátach nemáme outliere, ktoré by mohli skresľovať výsledné odhady jednotkových cien. Rozhodli sme sa najprv analyzovať vhodnosť použitia 4 rôznych metód na naše dáta a následne ďalej pracovať len s tou najrelevantnejšou. V našej práci sme teda najprv analyzovali nasledujúce metódy na identifikáciu outlierov:

- Tukeyho metóda s konštantou 1,5,
- Tukeyho metóda s konštantou 3,0,
- metóda Adjusted Boxplot,
- metóda vychádzajúca z analýzy k nestabilite,

kde pod pojmom „Tukeyho metóda s konštantou 1,5“ sme v našej práci rozumeli vylúčenie každej hodnoty prevyšujúcej vnútorné hranice vysvetlené v kapitole 2.3 a pod pojmom „Tukeyho metóda s konštantou 3,0“ sme v našej práci rozumeli vylúčenie každej hodnoty prevyšujúcej vonkajšie hranice, taktiež vysvetlené v kapitole 2.3.

Metóda vychádzajúca z analýzy k nestabilite vychádzala z poznatkov popísaných v kapitole 3.5, t. j. ako outliere sme pre jednotkové ceny identifikovali hodnoty, ktoré sa od mediánu líšili o viac ako 12-násobok alebo o 50 tis. €, čo sú maximálne hranice, pri ktorých sme si mohli byť istý, že hodnota bola outlier. Nepoužili sme násobok ako $\sqrt{12}$, resp.

absolútnu hodnotu 25 tis. €, nakoľko sme nevedeli vopred predpokladať, čo by bol skutočný stred. Vedeli sme len, aký bol rozdiel medzi minimálnou a maximálnou hodnotou, takže pri nesymetrických dátach by prísnejšie hranice mohli nejaké korektné hodnoty identifikovať ako outliere na úkor ponechania skutočných outlierov. Metóda vychádzajúca z analýzy k nestabilite nám tak slúžila skôr len ako limit, koľko položiek možno považovať za jednoznačných outlierov, ale neidentifikuje všetkých, nakoľko outlier by mohol byť aj menší ako 12-násobok alebo 50 tis. €. Túto metódu sme preto v našej práci použili len ako benchmark ostatných metód. Ak by totiž iné metódy identifikovali menej outlierov, znamenalo by to, že nie sú vhodné pre naše dáta.

3.7 Bodové odhadovanie jednotkových cien

Na samotnú tvorbu bodových odhadov jednotkových cien sme analyzovali 4 rôzne metódy vychádzajúce z poznatkov popísaných v kapitole 2.4 a to:

- nevážený aritmetický priemer („aritmetický priemer“),
- priemerné ceny vážené množstvom (ďalej len ako „vážený priemer“),
- medián,
- optimálny percentil.

Pre metódu vážený priemer sme za váhy považovali nakúpené množstva danej položky, pričom s danou metódou malo zmysel sa zaoberať len za predpokladu existencie objemnej zľavy v dátach. Objemová zľava je bežný ekonomický jav, keď dodávatelia poskytujú zľavy pri nákupe väčších množstiev tovarov alebo služieb, čo môže viesť k negatívnej korelácii medzi množstvom a jednotkovou cenou položiek, t. j. čím je nakúpené väčšie množstvo danej položky, tým by mala byť cena nižšia. Myšlienku metódy vážený priemer bolo možné zapísať aj ako

$$\text{vážený priemer} = \frac{\sum \text{celkových cien danej položky}}{\sum \text{nakúpeného množstva danej položky}}. \quad (3.1)$$

Metóda optimálny percentil bola nami navrhnutá metóda vychádzajúca z idey modusu v dátach. Rozhodli sme preskúmať nielen medián jednotkových cien položky (t. j.

50. percentil), ale aj ostatné percentily, nakoľko ak by dáta neboli symetrické, tak centrita dát by bola posunutá. Teda skutočnú hodnotu by najlepšie reprezentovala hodnota, okolo ktorej sa dáta najviac zhľukujú.

Spôsob výpočtu metódy optimálneho percentilu sme si definovali nasledovne:

1. Usporiadajú sa dáta do vzostupného poradia.
2. Pre každý percentil sa vypočíta hodnota indexu h ako

$$h = (n - 1) \cdot p + 1,$$

kde n je počet dátových bodov a p je hodnota z intervalu $[0;1]$, ktorá zodpovedá danému percentilu dát.

3. Rozdelí sa index h na celú časť k a desatinnú časť d ako

$$k = \lfloor h \rfloor,$$

$$d = h - k.$$

4. Vypočíta sa hodnota percentilu Q na základe hodnôt v dátovom súbore tak, že

- ak index h je celé číslo, potom je hodnota percentilu Q priamo zodpovedajúca hodnota dát na indexe k ,
- inak ak index h nie je celé číslo, potom je hodnota percentilu Q lineárna interpolácia medzi hodnotami na indexoch k a $k + 1$. Lineárna interpolácia sa vypočíta ako

$$Q = data[k] + d \cdot (data[k + 1] - data[k]),$$

kde $data[k]$ a $data[k + 1]$ sú hodnoty v usporiadanom dátovom súbore na indexoch k a $k + 1$.

5. Následne pomocou vopred vybranej metódy na validovanie odhadu v kombinácii s vopred vybranou metódou na meranie presnosti odhadu sa vyberie ten percentil, ktorý by vykazoval najlepšie výsledky. Tento vybraný percentil sme potom považovali za optimálny percentil.

3.8 Porovnávanie metód odhadovania jednotkových cien

Pre validovanie odhadu pri porovnávaní metód, ktorými sme odhadovali jednotkové ceny, sme sa rozhodli použiť krížovú validáciu, nakoľko je v štatistike považovaná za vhodnejšiu metódu ako obyčajná validácia z viacerých dôvodov. Po prvé, umožňuje poskytnúť odhad výkonnosti modelu na nových dátach, čo je dôležitý aspekt pri hodnotení jeho presnosti a robustnosti. Po druhé, môže pomôcť zabrániť nadmernému prispôbaniu modelu trénovacej dátovej sady, čo by mohlo viesť k nadhodnoteniu presnosti modelu na neznámych dátach. Po tretie, umožňuje využiť celú dostupnú dátovú sadu na trénovanie a validáciu, čím zvyšuje spoľahlivosť výsledkov.

Krížovú validáciu sme vykonali metódou náhodného rozdelenia. Túto metódu sme si vybrali z viacerých dôvodov:

- Flexibilita – Umožňuje flexibilné nastavenie veľkosti trénovacej a validačnej dátovej sady podľa potreby. To znamená, že metódu je možné prispôbiť rôznym požiadavkám na úroveň presnosti, rýchlosť alebo zdroje potrebné na vykonanie krížovej validácie.
- Robustnosť – Je schopná odolať viacerým problémom spojených s dátami, akými sú nevyváženosť tried, zhlukovanie dát alebo závislosť pozorovaní. Náhodné rozdelenie dát do trénovacích a validačných dátových sád zabezpečuje, že model je trénovaný na rôznorodej množine príkladov, čo zvyšuje jeho schopnosť generalizovať na nové dáta.
- Rýchlosť – Nevyžaduje komplexné algoritmy na rozdelenie dát alebo opakované prehodnotenie modelu, môže byť tak výpočtovo rýchlejšia ako niektoré iné metódy krížovej validácie.
- Náhodná variabilita – Zohľadňuje náhodnú variabilitu v dátach tým, že opakovane rozdeľuje dáta na trénovacie a validačné dátové sady. Týmto spôsobom môže táto metóda poskytnúť lepší odhad výkonnosti modelu ako metódy, ktoré pracujú iba s jedným pevným rozdelením dát, čo znižuje riziko, že výsledky validácie budú ovplyvnené špecifickým rozdelením dát.

V literatúre neexistuje jednotný konsenzus ohľadom optimálneho počtu opakovaní pri

krížovej validácii metódou náhodného rozdelenia. Počet opakovaní môže závisieť od rôznych faktorov, akými sú veľkosť dátového súboru, zložitosť modelu, čas potrebný na tréning a validovanie a tiež od odhadovanej variability výkonnosti modelu. Všeobecne platí, že vyšší počet opakovaní zvyšuje presnosť odhadu výkonnosti modelu, ale zároveň zvyšuje aj výpočtovú náročnosť krížovej validácie. Niektoré štúdie napr. [1, 8, 14] naznačujú, že pre väčšinu úloh v strojovom učení je dostatočný počet opakovaní v rozmedzí 5 až 10. Pre náš prípad sme sa rozhodli pre hornú hranicu 10 opakovaní. Viac sme neuvažovali z dôvodu vysokej výpočtovej náročnosti.

Nakoľko aj tie najväčšie štandardné železničné investičné projekty majú do 30 tis. položiek v rozpočte (neplatí len pre investičné projekty železničných uzlov, ktoré ale nie sú štandardnými projektami, nakoľko ich výskyt je zriedkavý[23]), sme považovali za nutné, aby validačná dátová sada mala aspoň 30 tis. položiek. Tomu by malo zodpovedať pomer rozdelenia trénovacej databázy v pomere 85 % ku 15 % v prospech trénovacej dátovej sady. Takýto pomer sme považovali za postačujúci, nakoľko aj v literatúre napr. v [8] sa bežne tento pomer definuje medzi 80 % ku 20 % až 90 % ku 10 %.

Postup krokov krížovej validácie sme vykonávali nasledovne:

1. Očistili sme trénovaciu databázu o nerelevantné položky popísaných v kapitole 3.2.
2. Náhodne sme rozdelili dáta v trénovacej databáze na 85 % pre trénovaciu dátovú sadu a 15 % pre validačnú dátovú sadu.
3. Vykonali sme štandardizáciu váh projektov popísanú v kapitole 3.3.
4. Identifikovali sme nestabilné položky v trénovacej dátovej sade podľa postupu popísaného v kapitole 3.5.
5. Detegovali sme outlierov medzi unikátnymi naindexovanými jednotkovými cenami naprieč všetkými projektami s vybranou najvhodnejšou metódou na identifikáciu outlierov a následne sme o nich očistili naše dáta.
6. Vypočítali sme odhady jednotkových cien jednotlivými relevantnými metódami na bodové odhadovanie jednotkových cien popísaných v kapitole 3.7.
7. Benchmarkovali sme validačnú dátovú sadu týmito nájdenými odhadmi jednotkových cien.

8. Vypočítali sme presnosť odhadu vhodnou metódou na meranie presnosti odhadov. Za vhodnú metódu na meranie presnosti odhadov sme si zvolili metriku MAPE, nakoľko sme považovali metriky založené na meraní absolútnych chýb za nekorektné používať v prípade cenovo dynamicky rozdielnom rozpočte, ktorý metóda náhodného rozdelenia pri krížovej validácii generuje.
9. Opakovali sme 2. až 8. krok 10-krát.
10. Vypočítali sme priemernú presnosť odhadu vhodnou metódou na meranie presnosti odhadov zo všetkých iterácií pre všetky použité metódy a porovnali.

3.9 Testovanie modelu na dátach simulujúcich reálne podmienky v praxi

Náš navrhnutý model odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov sme sa rozhodli ešte otestovať na testovacej dátovej sade simulujúcej reálne podmienky v praxi, ako sme avizovali už v kapitole 3.1. Rozhodli sme sa na nej výsledky nášho modelu konfrontovať s ocenenými výkazmi výmer daných rozpočtov. Taktiež so súčasným spôsobom benchmarkovania jednotkových cien stavebných prác a materiálov Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI, či skutočne prichádza k zlepšeniu súčasného stavu. Ako aj s odhadom jednotkových cien stavebných prác a materiálov komerčnou cenovou databázou CENEKON, nakoľko ide o najpoužívanejšiu komerčnú cenovú databázu pre stavebné projekty na slovenskom trhu. Pri meraní presnosti sme za skutočné hodnoty považovali zmluvné ceny v rozpočtoch, čo bol silný predpoklad, nakoľko ceny v rozpočtoch nemusia zodpovedať skutočnej trhovej cene, napr. z dôvodu nízkej účasti vo verejnom obstarávaní.

Porovnávanie jednotlivých spôsobov odhadovania jednotkových cien stavebných prác a materiálov sme sa rozhodli vykonať viacerými spôsobmi pre získanie komplexnejšieho pohľadu. Nakoľko každý z daných spôsobov odhadovania jednotkových cien benchmarkuje iné položky, bolo by nekorektné ich porovnávať metódami na základe absolútnych chýb. Preto sme sa aj tu rozhodli pre aplikovanie metriky MAPE. Najprv sme jednotlivé spôsoby odhadovania jednotkových cien porovnali s meraním presnosti metrikou MAPE agregovane za projekty, čím sme získali pohľad na projekty rovnocenne.

Nakoľko však testovacia dátová sada obsahovala aj špecifické projekty, ktoré sa v praxi vyskytujú ojedinele, za hlavný ukazovateľ sme si zobrali meranie presnosti metrikou MAPE za celú testovaciu dátovú sadu ako celok, čo je v podstate len absolútna percentuálna odchýlka (angl. *absolute percentage error*) (APE). Týmto spôsobom sme získali ucelený pohľad za určité časové obdobie v praxi.

Nakoniec sme sa rozhodli aplikovať aj váženú MAPE, t. j. wMAPE. Keďže sme mohli očakávať, že čím menšiu časť projektu je daný spôsob odhadovania jednotkových cien schopný benchmarkovať, tým vyššia môže byť chybovosť odhadov, rozhodli sme sa priradiť váhy v metrike wMAPE na základe benchmarkovaného objemu. Prvým dôvodom bolo zistiť, ako si jednotlivé spôsoby odhadovania jednotkových cien viedli medzi sebou v projektoch, kde majú vysoký benchmarkovaný objem, a teda bolo očakávané, že pre dané špecifikum projektu mali aj dostatok dát a naopak, aby sme nedávali rovnakú váhu hodnotám, ktoré vznikli z nedostatku dát. Druhým dôvodom bolo odvolanie sa na reprezentatívnu vlastnosť našej cenovej databázy, v ktorej sa nachádzajú takmer všetky významné železničné investičné projekty od roku 2012. Metrikou wMAPE na základe benchmarkovaného objemu sme sa tak snažili zohľadniť silné stránky jednotlivých spôsobov odhadovania jednotkových cien.

Pri hodnotení efektívnosti nášho modelu sme sa zamerali aj na *benchmarkovaný objem* vyjadrený v percentách. Tento ukazovateľ kvantifikuje podiel rozpočtu, ktorý sme boli schopní benchmarkovať našim modelom, t. j. našli sme aspoň jedno pozorovanie položky v našej cenovej databáze, ktorá nebola odfiltrovaná pri čistení dát alebo pri identifikovaní nestabilných položiek. Benchmarkovaný objem sme vypočítali ako

$$\text{benchmarkovaný objem} = \frac{\sum_{i=1}^m \text{celková zmluvná cena benchmarkovanej } i\text{-tej položky}}{\sum_{j=1}^n \text{celková zmluvná cena } j\text{-tej položky}},$$

kde n je počet všetkých položiek v dátovej sade, ktorú sme benchmarkovali, m je počet položiek v rovnakej dátovej sade, ktoré sme dokázali benchmarknúť, pričom platí $m \leq n$.

Okrem pohľadu na konkrétne projekty, či testovaciu dátovú sadu ako celok, sme sa na presnosť nášho modelu pozreli aj z pohľadu relevantných typov profesijných odborov, kde sme taktiež použili metriku MAPE.

Na záver sme vykonali ešte analýzu robustnosti našej cenovej databázy, t. j. zaujímalo nás, ako sa mení presnosť cenovej databázy na testovacej dátovej sade pri postupnom pridávaní projektov. Cieľom bolo preveriť, či zistená presnosť nášho modelu nebola len

zhoda náhod, ale ide o systematické zlepšenie presnosti pri rozširovaní cenovej databázy z dôvodu väčšieho počtu dát, a teda zlepšujúcej sa aproximácie ku skutočnej hodnote daných položiek. Preto sme vykonali stochastickú simuláciu rozširovania databázy pomocou metriky RMSE, resp. MAE. Keďže pridaná hodnota projektu pre cenovú databázu závisí len od predošlých projektov, rozhodli sme sa pre stochastický prístup pridávania projektov s opakovaním cyklu. Projekty sme pridávali po troch kvôli vysokej výpočtovo-časovej náročnosti, ak by sme ich pridávali po jednom. Proces sme opakovali 10-krát, aby sme dosiahli hodnovernú reprezentáciu procesu rozširovania našej cenovej databázy.

Postup krokov merania robustnosti našej cenovej databázy sme vykonávali nasledovne:

1. Vypočítali sme RMSE, resp. MAE celkovej odhadovanej ceny testovacej dátovej sady určenej projektantmi ku celkovej zmluvnej hodnote testovacej dátovej sady.
2. Vybrali sme všetky projekty z trénovacej databázy.
3. Projekty sme náhodne zoradili do poradia od 1 po 58.
4. Vybrali sme prvé 3 projekty z poradia, ktoré sme použili ako vstupné dáta do nášho modelu.
5. Vypočítali sme odhady jednotkových cien našim modelom.
6. Benchmarkovali sme testovaciu dátovú sadu týmito nájdenými odhadmi jednotkových cien s tým, že ak sme nemali danú porovnávanú položku v databáze, priradili sme jednotkovú cenu od projektantov. Takto sme mohli pozorovať, ako sa náš model vyvíja postupným odpútaním sa od jednotkových cien určených projektantmi.
7. Vypočítali sme presnosť odhadu metrikou RMSE, resp. MAE.
8. Vstupné dáta do nášho modelu sme rozšírili o ďalšie 3 projekty v poradí.
9. Kroky 5 až 8 sme opakovali pokým vstupné dáta do nášho modelu neobsahovali všetkých 58 projektov z trénovacej databázy.
10. Opakovali sme všetky predošlé kroky 10-krát.
11. Z týchto 10 opakovaní sme vypočítali priemerný vývoj robustnosti a grafickým znázornením zistili, či bol skutočne klesajúci trend.

4 Štatistická analýza a výsledky

V poslednej kapitole predloženej diplomovej práce sme sa podrobne zaoberali samotnou analýzou našich dát a jej výsledkami. Po úvodnom preskúmaní rôznych systémov indexácie, sme sa prostredníctvom vizualizácie v podobe histogramu a koeficientu vychýlenosti venovali kvantifikácii vychýlenosti dát, nakoľko bola potrebná pre ďalšie analýzy. Následne sme vykonali analýzu výkonnosti metód identifikujúcich outlierov a analyzovali sme aj hypotézu o existencii objemnej zľavy z dôvodu potvrdenia alebo vylúčenia metódy váženého priemeru na odhadovanie jednotkových cien. Vychádzajúc z poznatkov o našich dátach sme ešte kalibrovali metódu optimálny percentil a nakoniec sme krížovou validáciou porovnali relevantné metódy odhadovania jednotkových cien.

Záver kapitoly sme venovali výsledkom benchmarkovania testovacej dátovej sady navrhnutým modelom a výsledky porovnania s ocenenými výkazmi výmer tejto dátovej sady od projektantov, súčasným spôsobom benchmarkovania jednotkových cien stavebných prác a materiálov tejto dátovej sady Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI, ako aj s odhadom jednotkových cien stavebných prác a materiálov tejto dátovej sady komerčnou cenovou databázou CENEKON.

4.1 Analýza systémov indexácie

Pred akoukoľvek analýzou dát bolo potrebné najprv rozhodnúť, ktorý systém indexácie jednotkových cien by bol vhodnejší pre náš typ dát. Porovnávali sme tri rôzne systémy indexácie, a to systém kódov odborov, metodický pokyn č. 19/2022 a náš systém názvov objektov. Výber najvhodnejšej metódy pre odhadovanie cien by nemal byť ovplyvnený systémom indexácie, nakoľko zmyslom indexácie nie je spresniť odhadovanie jednotkových cien, ale dať dáta na spoločnú cenovú úroveň pre zabezpečenie ich porovnateľnosti. Vhodný systém indexovania by mal preto zosúladiť ceny z rôznych období na čo najpodobnejšie hodnoty s minimálnou mierou fluktuácie jednotkových cien položiek po indexácii. Ak by boli staršie projekty po indexácii systematicky nadhodnocované alebo podhodnocované v porovnaní s novšími projektmi, problém by spočíval skôr v nesprávnej indexácii, než v ich jednotkových cenách položiek.

Na základe vyššie uvedeného sme preto analyzovali zmeny vo výberových variančných

koeficientoch. Čím by bol nižší úhrnný výberový variančný koeficient s_{sum} , ktorý sme si definovali ako

$$s_{\text{sum}} = \sum_{i=1}^n \frac{s_i}{\bar{x}_i},$$

kde n reprezentuje celkový počet položiek, s vyjadruje výberovú štandardnú odchýlku pre i -tu položku a $\bar{x}_i$ vyjadruje aritmetický priemer pre i -tu položku, tým daný systém indexácie možno považovať za vhodnejší.

Indexovanie podľa metodického pokynu č. 19/2022 však aplikuje na dáta iba koeficient zmeny, kým nami navrhnutý systém názvov objektov aplikuje na dáta 3 rôzne indexy a v systéme kódov odborov aplikuje na dáta až 5 rôznych indexov. Čím vyšší je počet použitých indexov, tým prirodzenejšie rastie miera fluktuácie jednotkovej ceny položky, keďže položka sa môže indexovať rôznymi indexmi v závislosti od objektu, na ktorom sa nachádza. Aby sme vôbec mohli porovnať úroveň úhrnného výberového variančného koeficientu medzi rôznymi systémami indexácie, ktoré používajú rôzny počet indexov, kategorizovali sme položky v cenovej databáze na základe kódu položky, mernej jednotky, kódu indexu priradeného systémom kódov odborov a kódu indexu priradeného naším systémom názvov objektov. Týmto spôsobom sme štandardizovali systémy indexácie medzi jednotlivými metódami. Následne sme vypočítali výberové variančné koeficienty pre všetky kategorizované položky po preindexovaní jednotkových cien do spoločnej cenovej úrovne 4. štvrtroka 2022 jednotlivými systémami indexácie a tieto výberové variančné koeficienty sme sčítali, aby sme získali úhrnný výberový variančný koeficient. Výsledky tejto analýzy sme uviedli do tab. 4.1.

Tab. 4.1: Výsledky kategorizovaného úhrnného výberového variančného koeficientu pre rôzne systémy indexácie

Systém kódov odborov	MP č. 19/2022	Systém názvov objektov
$s_{\text{sum}} = 7\,597,65$	$s_{\text{sum}} = 7\,609,89$	$s_{\text{sum}} = 7\,599,47$

Zdroj: vlastné spracovanie

Z našej analýzy vyplynulo, že systémy boli porovnateľné medzi sebou, pričom indexácia podľa metodického pokynu č. 19/2022 pre potreby indexovania jednotkových cien sa ukázala ako najmenej vhodná. Tento výsledok sme očakávali, nakoľko týmto systémom sme použili iba jeden koeficient pre indexovanie všetkých položiek. Systém kódov odborov

a náš systém názvov objektov dosiahli podobné výsledky, čo bolo hlavne dôsledkom toho, že väčšina položiek v oboch systémoch je indexovaná rovnako. Na základe systému kódov odborov sa totiž indexovalo:

- 60,66 % cenovej databázy podľa 212 Železnice a dráhy,
- 25,88 % cenovej databázy podľa 214 Mosty, nadjazdy, tunely a podzemné dráhy,
- 6,75 % cenovej databázy podľa 124 Budovy pre dopravu a telekomunikáciu,
- 4,69 % cenovej databázy podľa 211 Cestné a miestne komunikácie,
- 2,02 % cenovej databázy podľa 222 Miestne potrubné a káblové rozvody,

kým podľa nášho systému názvov objektov sa indexovalo:

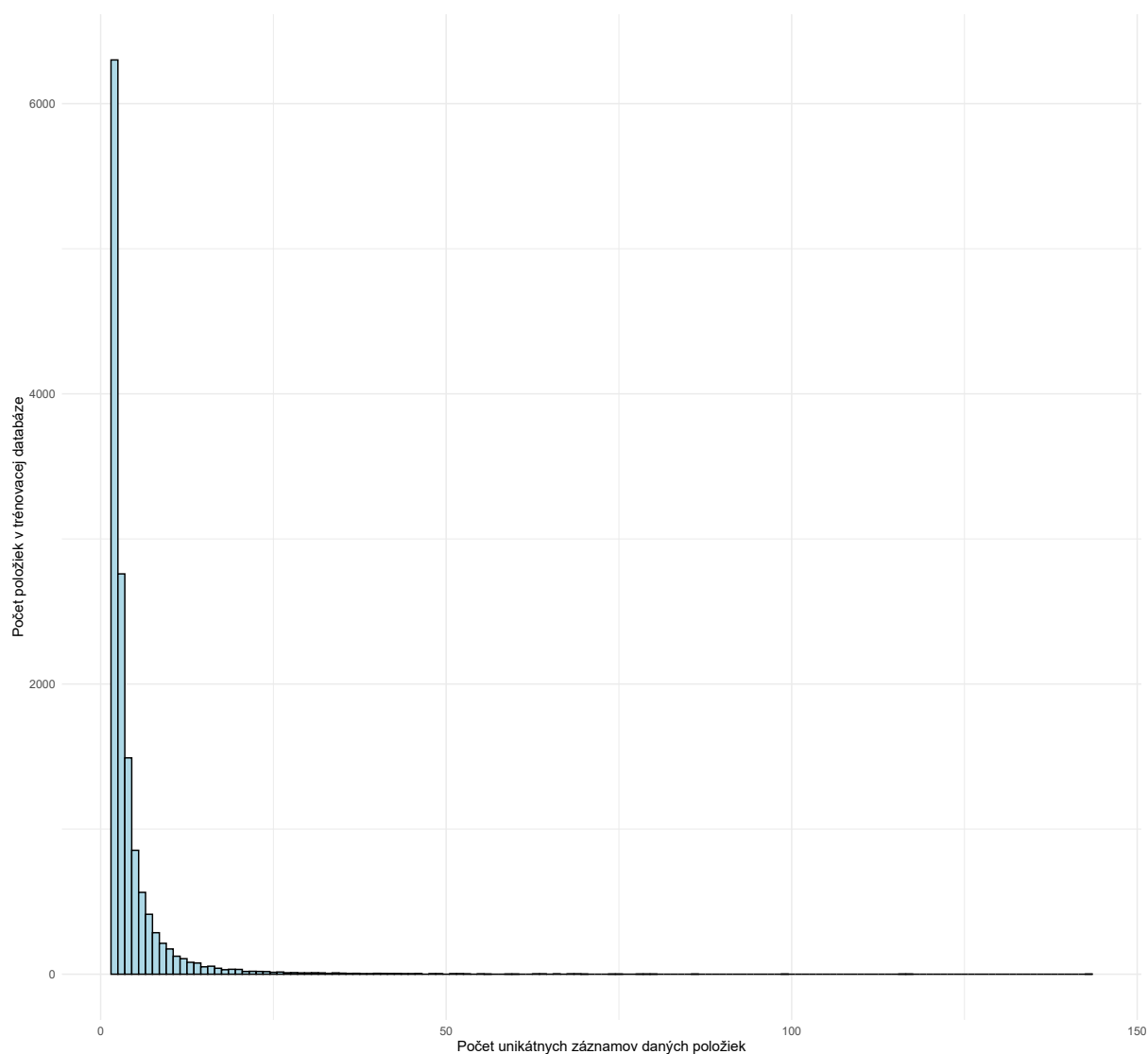
- 71,90 % cenovej databázy podľa 212 Železnice a dráhy,
- 26,98 % cenovej databázy podľa 214 Mosty, nadjazdy, tunely a podzemné dráhy,
- 1,11 % cenovej databázy podľa 124 Budovy pre dopravu a telekomunikáciu.

Rozdiel medzi týmito dvoma systémami bol príliš malý na to, aby jednoznačne preukázal, ktorý z daných dvoch systémov indexácie by mal byť vhodnejší pre naše dáta, keďže rozdiel mohol byť spôsobený len obmedzeniami našej cenovej databázy. Preto sme z praktických dôvodov uprednostnili náš systém, keďže nie je závislý na nepovinných údajoch v rozpočtoch, ako je tomu v prípade systému kódov odborov a vyžaduje menej indexov.

4.2 Analýza charakteru dát

Keď už sme si vybrali najvhodnejší systém indexácie jednotkových cien, pred ďalšou analýzou sme sa zamerali na charakter našich dát, s ktorými sme pracovali v našej tréningovej databáze. Naším cieľom bolo preskúmať analýzu unikátnych záznamov položiek a ich početnosti, čo nám poskytlo dôležité informácie o charakteristike našich dát, ktoré sme mohli využiť na navrhnutie a úpravu metód analýzy pre presnejšie odhady jednotkových cien, ako aj skúmanie početnosti položiek podľa kódov profesijných odborov, ktoré nám zas poskytlo možnosť identifikovať profesijné odbory, s ktorými bolo pre nás v ďalších analýzach relevantné sa zaoberať.

Najprv sme vykonali analýzu rozloženia početnosti unikátnych záznamov položiek v rámci trénovacej databázy. Pojem „unikátny záznam položky“ sme definovali ako každý záznam položky, ktorý má v rámci projektu jedinečnú neindexovanú jednotkovú cenu. Poznanie tohto rozloženia malo význam pre výber vhodných metód napr. na identifikáciu outlierov. V rámci tejto analýzy sme sa zamerali na položky s aspoň 2 záznamami, pretože položky s nižším počtom záznamov (teda len s jedným) neboli relevantné pre našu analýzu ani výber vhodných metód. Na základe týchto kritérií sme zostavili stĺpcový graf znázorňovaný na obr. 4.1, z ktorého je zrejmé, že prevládali položky s nízkym počtom unikátnych záznamov.



Obr. 4.1: Početnosť unikátnych záznamov položiek v trénovacej databáze

Zdroj: vlastné spracovanie

Zaujímavým aspektom bola aj analýza frekvencie kódov profesijných odborov v našej trénovacej databáze obsahujúca väčšinu železničných investičných projektov, ktorých zmluvy boli uzavreté v rokoch 2012 až 2021 vrátane takmer všetkých veľkých projektov. Vzhľadom na široké zastúpenie projektov sme považovali naše dáta za dostatočne reprezentatívne. Výsledky tejto analýzy sú uvedené v tab. 4.2.

Tab. 4.2: Rozdelenie kódov profesijných odborov v trénovacej databáze

Kód odboru	Indexovaná zmluvná cena bez DPH	Percentuálny podiel
21	239 632 497,00 €	14,17 %
22	73 990 294,35 €	4,38 %
23	2 153 722,36 €	0,13 %
24	33 035 281,93 €	1,95 %
25	30 651 273,14 €	1,81 %
26	4 048 018,91 €	0,24 %
27	6 527 550,53 €	0,37 %
28	4 026 241,54 €	0,24 %
29	6 539 169,00 €	0,39 %
31	7 470 971,06 €	0,44 %
32	459 071 125,10 €	27,14 %
33	392 742 019,20 €	23,22 %
34	109 713 815,40 €	6,49 %
35	209 661 132,30 €	12,40 %
36	12 397 326,73 €	0,73 %
37	21 657 025,17 €	1,28 %
38	66 398 943,48 €	3,93 %
39	11 655 725,50 €	0,69 %

Zdroj: vlastné spracovanie

Výsledky v tab. 4.2 ukazujú najväčšie zastúpenie nasledujúcich kódov profesijných odborov:

- 21 – Železničné zabezpečovacie zariadenia,

- 32 – Železničný zvršok a spodok, nástupištia, káblová chráničková trasa, priepusty,
- 33 – Mosty a umelé stavby,
- 35 – Trakčné vedenie a energetika.

Ostatné profesijné odbory boli zastúpené v miere menšej ako 10 % a dalo sa predpokladať, že v nich by sme mali mať nízky benchmarkovaný objem, resp. nízku presnosť odhadovaných jednotkových cien položiek bežne spadajúce pod tieto objekty z dôvodu nízkeho počtu dát. Preto v ďalších analýzach sme sa primárne zamerali na typy objektov spadajúce do odborov s najväčším zastúpením.

4.3 Analýza vychýlenosti dát

Pred samotnou analýzou metód na identifikáciu outlierov v dátach, ako aj metód na odhadovanie jednotkových cien, bolo potrebné ešte vykonať analýzu vychýlenosti našich dát. Tá nám pomohla práve pri výbere vhodnej metódy na identifikáciu outlierov, ako aj na zúženie intervalu, v ktorom by sa mal nachádzať optimálny percentil pre odhadovanie jednotkových cien položiek.

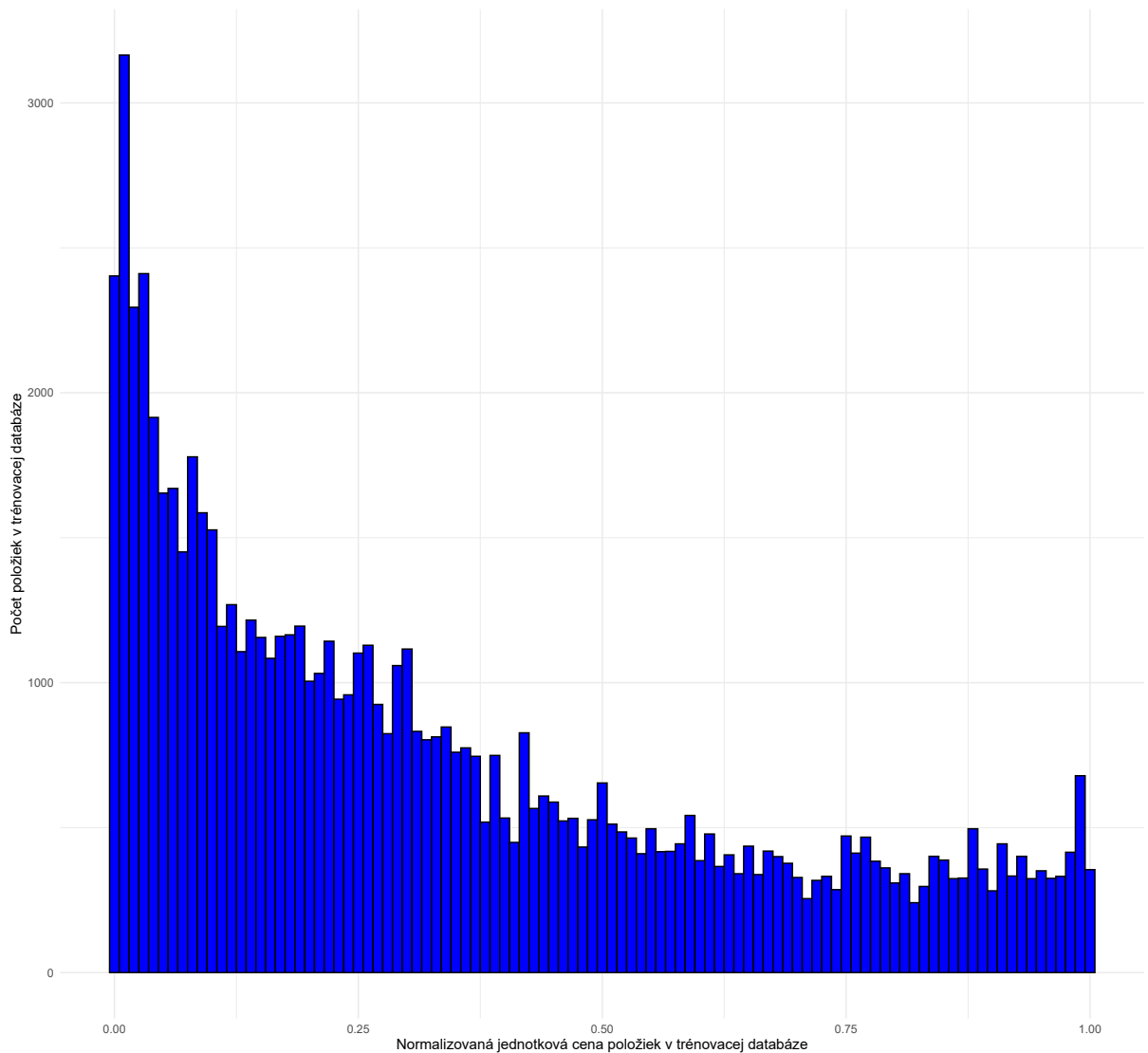
Rozhodli sme sa najprv vizualizovať dáta prostredníctvom histogramu, čo nám umožnilo lepšie pochopiť tvar rozdelenia našich dát. Aby sme mohli porovnať jednotkové ceny položiek, ktoré sa od seba odlišujú, normalizovali sme si jednotkové ceny do intervalu $[0; 1]$. Vychádzali sme z bežne v štatistike používanej metódy na normalizáciu dát tzv. min-max škálovanie[36], ktoré sa dá vyjadriť ako

$$\text{normalizovaná cena}_i = \frac{\text{cena}_i - \text{cena}_{\min}}{\text{cena}_{\max} - \text{cena}_{\min}}, \quad (4.1)$$

kde normalizovaná cena_i je normalizovaná indexovaná jednotková cena pre i -tu položku, cena_i je pôvodná indexovaná jednotková cena pre i -tu položku, $\text{cena}_{\min}$ je minimálna indexovaná jednotková cena zo všetkých položiek a $\text{cena}_{\max}$ je maximálna indexovaná jednotková cena zo všetkých položiek.

Vzorec (4.1) umožňuje previesť jednotkové ceny na normalizovanú škálu, čo zjednodušuje porovnávanie jednotkových cien medzi rôznymi položkami. Z normalizovaných jednotkových cien sme následne odstránili položky, ktoré mali nulovú alebo jednotkovú hodnotu, nakoľko by prítomnosť týchto krajných hodnôt skresľovala pohľad na graf, keďže

všetky položky s aspoň dvoma záznamami mali oba tieto extrémny. Výsledný histogram na obr. 4.2 znázorňuje pozitívnu vychýlenosť dát, čo znamenalo, že jednotlivé položky vykazovali viac hodnôt s relatívne nízkymi jednotkovými cenami a menej hodnôt s relatívne vysokými jednotkovými cenami.



Obr. 4.2: Histogram normalizovaných jednotkových cien v trénovacej databáze

Zdroj: vlastné spracovanie

Aby sme získali presnejšie informácie a overili naše pozorovanie, vypočítali sme aj koeficient vychýlenosti podľa vzorca (2.1) pre každú položku, ktorá mala viac ako 1 záznam po štandardizácii váh projektov na trénovacej databáze. Následne sme agregovali položky podľa charakteru vychýlenosti, čo nám umožnilo získať podrobný pohľad na vlastnosti dát, ktoré následne ovplyvnia výber metód a prístupov pre ich spracovanie a analýzu.

Výsledky agregácie položiek podľa charakteru vychýlenosti sme vypísali do tab. 4.3.

Tab. 4.3: Početnosť položiek podľa ich vychýlenosti

Vychýlenosť	Počet položiek	Percentuálny podiel
Negatívna	2 600	18,46 %
Symetrická	6 505	46,18 %
Positívna	4 981	35,36 %

Zdroj: vlastné spracovanie

Pre definitívne potvrdenie pozitívnej vychýlenosti našich dát sme si vypočítali aj úhrnnú mieru koeficientu vychýlenosti S_{sum} , ktorý sme si definovali ako

$$S_{sum} = \sum_{i=1}^n \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s^3},$$

kde n reprezentuje celkový počet položiek, x_i je i -tá hodnota, $\bar{x}_i$ vyjadruje aritmetický priemer pre i -tu položku a s je výberová štandardná odchýlka hodnôt pre i -tu položku. Táto úhrnná miera koeficientu vychýlenosti vyšla v sume 2 658,92, teda kladný výsledok nám potvrdil pozitívnu vychýlenosť našich dát.

4.4 Analýza metód identifikujúcich outliere

Ešte pred samotným odhadovaním jednotkových cien bolo treba odfiltrovať outliere z dát. Najprv sme vykonali porovnanie, koľko outlierov identifikuje aká metóda po štandardizácii váh projektov na tréningovej databáze. Výsledky tohto porovnania sme spracovali do tab. 4.4.

Tab. 4.4: Výsledky porovnania metód identifikujúcich outliere

Použitá metóda	Počet identifikovaných outlierov
Tukeyho metóda s konštantou 1,5	3 070
Tukeyho metóda s konštantou 3,0	1 218
Metóda Adjusted Boxplot	3 154
Metóda vychádzajúca z analýzy k nestabilite	825

Zdroj: vlastné spracovanie

Ako sme už uviedli v kapitole 3.6, metódu vychádzajúcu z analýzy k nestabilite sme použili len ako benchmark ostatných metód, nakoľko identifikuje počet jednoznačných outlierov, ale neidentifikuje všetkých outlierov. Z porovnania metód identifikujúce outlierie vyplynulo, že všetky metódy identifikovali viac outlierov, ako metóda vychádzajúca z analýzy k nestabilite, a teda nebolo možné žiadnu z nich vylúčiť len na základe nedostatočného identifikovania počtu outlierov.

Z porovnania však tiež vyplynulo, že Tukeyho metóda s konštantou 1,5 a metóda Adjusted Boxplot boli pre naše dáta výkonovo veľmi podobné a podľa teórie popísanej v kapitole 2.3 je metóda Adjusted Boxplot vylepšenou Tukeyho metódou s konštantou 1,5, resp. v prípade symetrickosti dát by boli totožné. Nakoľko na základe analýzy vychýlenosti dát uvedenej v kapitole 4.3 sme potvrdili vychýlenosť našich dát, nemalo zmysel sa ďalej zaoberať Tukeyho metódou s konštantou 1,5.

Taktiež podľa výsledkov analýzy rozloženia početnosti unikátnych záznamov položiek znázornených na obr. 4.1 na str. 52, ktorou sme ukázali, že naše dáta majú nízku unikátnu početnosť záznamov, sa ani Tukeyho metóda s konštantou 3,0 v zmysle teórie popísanej v kapitole 3.6 nejavila ako vhodná metóda pre naše dáta a rozhodli sme sa preto ďalej používať už iba metódu Adjusted Boxplot.

4.5 Analýza existencie objemnej zľavy

Nakoľko metódu vážený priemer na odhadovanie jednotkových cien malo zmysel zvažovať len v prípade existencie objemnej zľavy v našich dátach, bolo potrebné pred samotným porovnávaním metód odhadovania jednotkových cien aj preskúmať možnú existenciu objemnej zľavy, t. j. či existovala štatisticky významná negatívna korelácia medzi množstvom a jednotkovou cenou v našich dátach.

Existenciu objemnej zľavy bolo možné preukázať prostredníctvom vzťahu medzi váženým priemerom a aritmetickým priemerom. Ak by totiž existovala objemná zľava v našich dátach, potom možno predpokladať, že čím väčšie je nakúpené množstvo nejakej položky, tým nižšia je jej jednotková cena a súčasne je vyššia jej váha podľa vzorca (3.1). Potom by vážený priemer musel byť nižší ako aritmetický priemer, resp. by malo platiť:

1. Ak by signifikantná väčšina položiek mala vážený priemer nižší ako aritmetický priemer, t. j. rozdiel váženého priemeru a aritmetického priemeru by bol záporný,

v dátach by bola možná existencia objemnej zľavy.

2. Ak by signifikantná väčšina položiek mala vážený priemer rovný aritmetickému priemeru, t. j. rozdiel váženého priemeru a aritmetického priemeru by bol nulový, cena by nezávisela od množstva.
3. Ak by signifikantná väčšina položiek mala vážený priemer vyšší ako aritmetický priemer, t. j. rozdiel váženého priemeru a aritmetického priemeru by bol kladný, v dátach by bola dokonca možná existencia objemnej prirážky (t. j. opak objemnej zľavy, keď s narastajúcim množstvom sa jednotková cena zvyšuje).
4. Ak by nenastala ani jedna vyššie zmienená situácia 1 až 3, t. j. neexistovala by signifikantná väčšina položiek, ktoré by jednoznačne preukázali významný vzťah medzi váženým priemerom a aritmetickým priemerom, potom by sa nedalo preukázať, že či cena závisí od množstva. V takom prípade, ak by počet položiek, ktoré by mali vážený priemer nižší ako aritmetický priemer, prevyšoval tie, ktoré by mali vážený priemer vyšší ako aritmetický priemer, by to signalizovalo sklon k objemnej zľave. Ak by to bolo opačne, tak by to signalizovalo sklon k objemnej prirážke.

Vytvorili sme tab. 4.5 zobrazujúcu počet položiek na základe vzťahu váženého priemeru k aritmetickému priemeru, pričom sme sa zamerali iba na položky zaznamenané v aspoň 2 projektoch, aby sme minimalizovali skreslenie výsledkov.

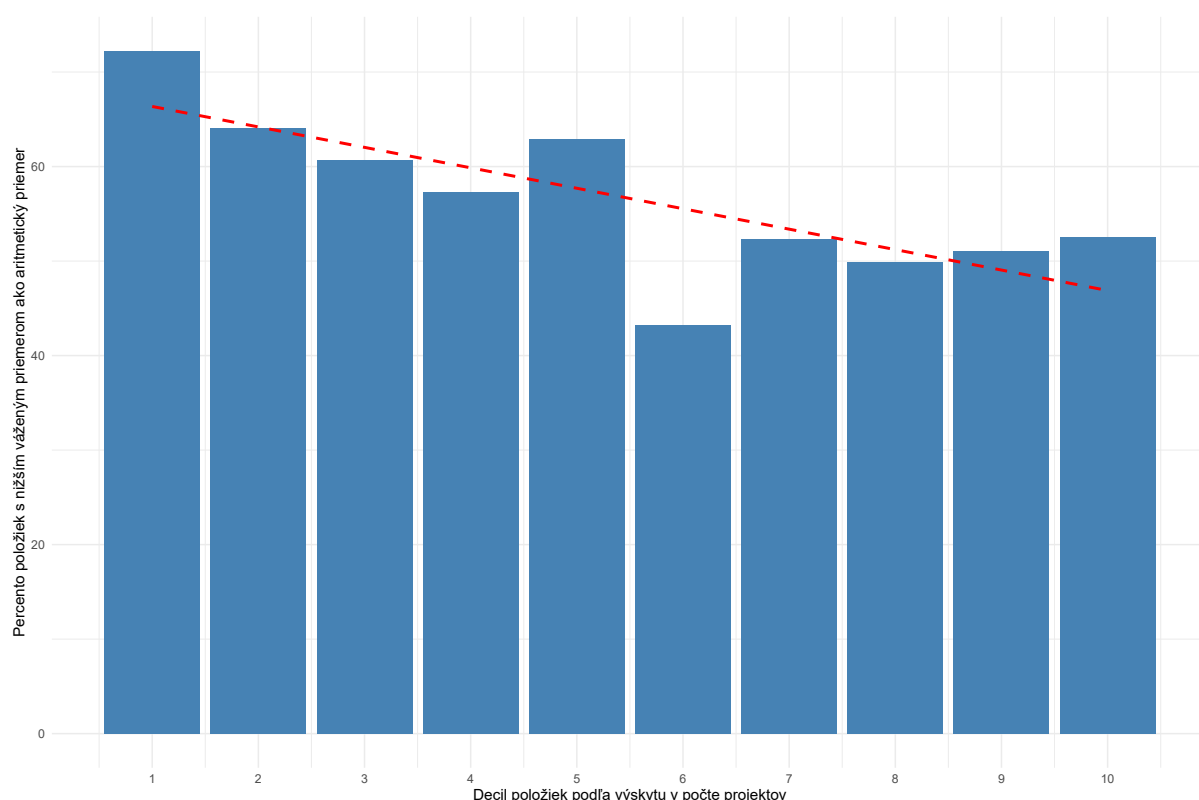
Tab. 4.5: Roztriedenie položiek podľa rozdielu váženého priemeru od aritmetického priemeru

Rozdiel	Počet položiek	Percentuálne zastúpenie
Záporný	5 814	51,46 %
Nulový	927	8,21 %
Kladný	4 556	40,33 %

Zdroj: vlastné spracovanie

Tab. 4.5 ukazuje, že len 51,46 % položiek malo nižší vážený priemer ako aritmetický priemer, existenciu objemnej zľavy tak jednoznačne preukázať nebolo možné. Dáta v trénovacej databáze však vykazovali sklon k objemnej zľave, nakoľko počet položiek, ktoré mali vážený priemer nižší ako aritmetický priemer, prevyšoval tie, ktoré mali vážený priemer vyšší ako aritmetický priemer.

Aby sme získali hlbšie pochopenie tohto sklonu, vytvorili sme stĺpcový graf na obr. 4.3, v ktorom každý stĺpec reprezentuje 10 % analyzovaných položiek zoradených vzostupne podľa početnosti ich záznamu vo viacerých projektoch. Prvý decil je tak 10 % položiek s najnižším záznamom v trénovacej databáze a posledný decil je zas 10 % položiek s najvyšším záznamom v trénovacej databáze. Opäť sme sa zamerali len na položky, ktoré boli zaznamenané v aspoň 2 projektoch, aby sme minimalizovali skreslenie výsledkov.



Obr. 4.3: Korelácia množstva a jednotkových cien položiek v trénovacej databáze

Zdroj: vlastné spracovanie

Trendová čiara lineárneho regresného modelu nachádzajúca sa na stĺpcovom grafe na obr. 4.3 ilustruje, že s narastajúcim počtom záznamov položiek vo viacerých projektoch klesá počet tých, ktoré majú nižší vážený priemer ako aritmetický priemer, až kým sa ustáli okolo 50 %. Inými slovami, s narastajúcim počtom záznamov je existencia objemnej zľavy menej preukázateľná.

Na základe výsledkov analýzy existencie objemnej zľavy nemala metóda váženého priemeru na odhadovanie jednotkových cien opodstatnenie pre naše dáta. Z tohto dôvodu sme túto metódu vylúčili z ďalšej analýzy.

4.6 Výsledky krížovej validácie

Nakoľko sme metódu váženého priemeru na odhadovanie jednotkových cien vylúčili analýzou existencie objemnej zľavy, samotnou krížovou validáciou sme porovnávali už len 3 metódy, a to aritmetický priemer, medián a optimálny percentil. Pred samotným porovnaním sme však museli najprv postupom popísaným v kapitole 3.7 vybrať na tréningovej dátovej sade percentil, ktorý by vykazoval najvyššiu priemernú presnosť na validačnej dátovej sade. Následne sme porovnali nami najoptimálnejší zvolený percentil s ostatnými metódami odhadovania jednotkových cien pomocou krížovej validácie.

4.6.1 Kalibrovanie metódy optimálny percentil

Keďže preskúmanie 100 percentilov 10 iteráciami krížovou validáciou na tak objemnej databáze by mohlo byť z časového hľadiska zdĺhavé a výpočtovo náročné, snažili sme sa hľadanie optimálneho percentilu z časového a výpočtového hľadiska ešte optimalizovať. Na základe vykonaných analýz sme si zúžili interval, v ktorom by sa mal optimálny percentil nachádzať. V kapitole 4.3 sme ukázali, že dáta v tréningovej databáze boli pozitívne vychýlené, preto preverovať percentily nad mediánom by nemalo opodstatnenie.

Aj keď obr. 4.2 na str. 55 vykazoval výraznú pozitívnu vyháňenosť, do histogramu neboli vzaté položky len s 2 záznamami, ktoré sú zo svojej podstaty symetrické, a teda celkovú vychýlenosť dát posúvajú smerom k mediánu. Vychádzajúc z tab. 4.3 na str. 56, ak by všetky pozitívne vychýlené položky boli vychýlené blízko k 0. percentilu a všetky negatívne vychýlené položky blízko k mediánu (t. j. 50. percentilu), čo by bol najextrémnejší prípad, tak priemerný percentil, ku ktorému by boli dáta vychýlené, by bol 32, keďže by platilo:

$$\frac{\lim_{i \rightarrow 0} 4981 \cdot i + 6505 \cdot 50 + \lim_{j \rightarrow 50} 2600 \cdot j}{4981 + 6505 + 2600} = \frac{4981 \cdot 0 + 6505 \cdot 50 + 2600 \cdot 50}{4981 + 6505 + 2600} \approx 32.$$

Ak by pozitívne vychýlené položky boli menej vychýlené, alebo negatívne vychýlené položky by boli viac vychýlené, ako vyššie zmienený príklad, priemerný percentil, ku ktorému by boli dáta vychýlené, by bol ešte bližšie k mediánu. Preto interval, v ktorom bolo dôvodné predpokladať, že by sa mohol nachádzať optimálny percentil, sme si určili ako [25; 50].

Kedže preskúmanie 26 percentilov by bolo stále dosť časovo zdĺhavé a výpočtovo náročné, rozhodli sme sa hľadať optimálny percentil v nami určenom intervale vybranou metódou voľnej optimalizácie, ktorou by sme neperspektívne percentily preskočili, a tým skrátili čas hľadania optimálneho percentilu. Inšpirovali sme sa metódou zlatého rezu, ktorá sa používa na hľadanie lokálneho minima pre numericky aproximované funkcie[30].

Pri hľadaní optimálneho percentilu bolo odôvodnené predpokladať konvexnú množinu pozorovaní v rámci hľadania optimálneho percentilu, pretože uchádzači o investičný projekt sa usilujú o získanie zákazky v rámci verejného obstarávania. To dosiahnu len ponukou nižšej ceny ako konkurenční uchádzači, najmä v prípade železničných projektov, kde je najnižšia cena zvyčajne jediným hodnotiacim kritériom[23]. V dôsledku takejto súťaže by sa vysúťažaná cena mala postupne približovať k „optimálnej cene“. To znamená, že bolo možné očakávať zoskupenie dát okolo nejakej optimálnej ceny.

Teda pre potreby hľadania globálneho minima, resp. optimálneho percentilu, inšpirujúc sa metódou zlatého rezu, bolo treba upraviť postup krížovej validácie pre tento účel:

1. Počiatočné hodnoty intervalu $[a, b]$, v ktorom sme chceli nájsť minimum funkcie $f(x)$, sme určili ako $a_0 = 25$ a $b_0 = 50$.
2. Potom sme vypočítali body x_1 a x_2 v intervale $[a, b]$ s použitím zlatého rezu

$$x_1 = a + (1 - \phi)(b - a),$$

$$x_2 = a + \phi(b - a),$$

kde ϕ je zlatý rez.

3. Následne nastáva postup krížovej validácie (spomenutý v kapitole 3.8), kde oba body x_1 a x_2 sa použili pre výpočet percentilu a zapísali sme ich priemernú presnosť na tých istých 10 náhodne vygenerovaných validačných dátových sadách. Hodnoty funkcie $f(x)$ sme vnímali ako priemernú presnosť počítajúc metrikou MAPE.
4. Určili sme minimum funkcie $f(x)$ na podintervale $[a, x_2]$ alebo $[x_1, b]$, a to:
 - ak $f(x_1) \leq f(x_2)$, potom sa minimum nachádza v intervale $[a, x_2]$,
 - ak $f(x_1) > f(x_2)$, potom sa minimum nachádza v intervale $[x_1, b]$.
5. Aktualizovali sme interval, ktorý zahŕňal už aj minimum.

6. Pokračovali sme v iteráciách, kým nebola dosiahnutá požadovaná presnosť, ktorú sme si stanovili ako $|a - b| < 0,5$ z dôvodu hľadania celočíselného percentilu.

Po 9 iteráciach cyklus skončil, pričom sme identifikovali hodnotu 38,17 ako optimálnu s priemernou hodnotou $\text{MAPE} = 2,66\%$ a výberovou štandardnou odchýlkou 2,91 %. Pre potreby našej práce sme ďalej použili najbližší celočíselný percentil, t. j. 38., ako optimálny percentil.

4.6.2 Porovnanie metód odhadovania jednotkových cien

Po nájdení optimálneho percentilu sme ho prostredníctvom krížovej validácie a metriky MAPE porovnali aj s metódami aritmetický priemer a medián. Na základe analýzy výsledkov krížovej validácie uvedených v tab. 4.6 na str. 63 je zrejmé, že metóda optimálneho percentilu dosiahla najnižšie hodnoty metriky MAPE, a teda aj najvyššiu presnosť. Metóda optimálneho percentilu dosiahla zároveň aj najnižšiu hodnotu výberovej štandardnej odchýlky, čo znamená, že bola aj najstabilnejšia z hľadiska presnosti. Metóda medián podľa očakávania skončila horšie, nakoľko už hľadanie optimálneho percentilu ho vylúčilo a metóda aritmetický priemer skončila najhoršie, čo sa tiež dalo očakávať, nakoľko pri pozitívne vychýlených dátach by mal byť aritmetický priemer väčší ako medián, a teda ešte viac vzdialený optimálnej hodnote.

4.7 Model odhadovania jednotkových cien

Výstupom našej analýzy bol návrh konkrétneho modelu odhadovania jednotkových cien stavebných prác a materiálov, ktorý je možné formulovať do nasledujúcich krokov:

1. Vstupné dáta sa očistia o nerelevantné položky popísaných v kapitole 3.2.
2. Následne sa vykoná štandardizácia váh projektov popísaná v kapitole 3.3.
3. Dáta sa ďalej očistia o nestabilné položky podľa postupu popísaného v kapitole 3.5.
4. Potom sa dáta očistia o outliere identifikované metódou Adjusted Boxplot popísanej v kapitole 2.3.
5. Na očistených dátach sa vypočítajú bodové odhady jednotkových cien 38. percentilom vypočítaným podľa postupu popísaného v kapitole 3.7.

Tab. 4.6: Porovnanie metód odhadovania jednotkových cien krížovou validáciou

Iterácia	Aritmetický priemer	Medián	Optimálny percentil
1.	11,09 %	6,42 %	0,83 %
2.	8,44 %	6,09 %	1,45 %
3.	18,48 %	14,58 %	7,08 %
4.	3,95 %	0,96 %	5,45 %
5.	16,24 %	10,90 %	3,75 %
6.	10,56 %	7,04 %	0,01 %
7.	13,21 %	10,44 %	2,60 %
8.	5,98 %	1,94 %	5,01 %
9.	14,79 %	11,61 %	4,67 %
10.	14,05 %	10,77 %	2,62 %
MAPE	11,68 %	8,07 %	3,35 %
Výberová štandardná odchýlka	4,58 %	4,37 %	2,24 %

Zdroj: vlastné spracovanie

4.8 Porovnanie s inými cenovými databázami

Po výbere najvhodnejšej metódy odhadovania jednotkových cien sme sa venovali implementácii a vyhodnoteniu výsledkov nášho modelu popísaného v kapitole 4.7 aplikáciou na dáta simulujúce reálne podmienky v praxi, pričom sme konfrontovali náš model aj s ocenenými výkazmi výmer od projektantov, so zaužívaným spôsobom benchmarkovania železničných investičných projektov Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI, ale aj najpoužívanejšou komerčnou cenovou databázou na slovenskom stavebnom trhu CENEKON. Oddelenie dopravných projektov ÚHP nám poskytlo ocenené výkazy výmer od projektantov a odhady jednotkových cien stavebných prác a materiálov cenovými databázami SFDI aj CENEKON pre našu testovaciu dátovú sadu[23]. Všetky spôsoby odhadovania jednotkových cien, ktoré sme porovnávali medzi sebou, sme previedli na spoločnú cenovú úroveň 4. štvrtroka 2022.

Podľa vyjadrení pracovníkov ÚHP, benchmarkovanie jednotkových cien testovacej dátovej sady pomocou cenových databáz SFDI trvalo 32 hodín čistej práce z dôvodov spomenutých v kapitole 1.2 a benchmarkovanie pomocou cenovej databázy CENEKON trvalo 2 hodiny čistej práce[23]. Naproti tomu, benchmarkovanie jednotkových cien našou tré-

novacou databázou trvalo len 37 sekúnd, čo jednoznačne možno považovať za výrazné zrýchlenie práce. V prílohe D sme uviedli podrobné výsledky benchmarkovania testovacej dátovej sady jednotlivými spôsobmi odhadovania jednotkových cien, a to ako podľa agregovaných projektov, tak aj podľa profesijných odborov. Hlavné výsledky sme vypísali do tab. 4.7 a podľa profesijných odborov do tab. 4.8 na str. 65.

Tab. 4.7: Hlavné výsledky benchmarku testovacej dátovej sady

	PHZ	Model	SFDI	CENEKON
Benchmarkovaný objem v €	406 409 094,14 €	236 146 586,61 €	209 632 988,28 €	142 602 802,61 €
Benchmarkovaný objem v %	100,00 %	58,11 %	51,58 %	35,09 %
MAPE agregovaných projektov	10,56 %	27,05 %	13,36 %	18,67 %
Benchmarkovaná časť rozpočtu APE	6,97 %	4,00 %	7,90 %	9,70 %
Vážené MAPE	10,56 %	12,60 %	12,28 %	28,90 %

Zdroj: vlastné spracovanie

Náš model dokázal benchmarknúť vyšší objem testovacej dátovej sady, ako cenové databázy SFDI alebo CENEKON a zároveň bol najbližšie ku celkovému zmluvnému rozpočtu testovacej dátovej sady použitou metrikou APE, čo bol náš hlavný ukazovateľ.

Z pohľadu metriky MAPE agregovane po rozpočtoch však náš model skončil najhoršie. Pri bližšom skúmaní príčiny tohoto javu sme zistili, že dôvodom boli špecifické projekty, u ktorých náš model zlyhával aj z pohľadu benchmarkovaného objemu (tab. D.1 v prílohe D), a to „Implementácia GSM-R do siete ŽSR, úsek Varín – Košice – Čierna nad Tisou štátna hranica“, kde náš model dosiahol iba 7,04 % benchmarkovaný objem a mal až 95,29 % relatívnu výchylku od zmluvnej ceny a „Rekonštrukcia Bratislavského tunela č. 2, ŽST Bratislava hl. st.“, kde náš model dosiahol 23,71 % benchmarkovaný objem a mal 79,75 % relatívnu výchylku od zmluvnej ceny. V našej trénovacej databáze chýbali projekty, ktoré by boli rovnako špecifické.

Obdobný projekt ku projektu „Implementácia GSM-R do siete ŽSR, úsek Varín – Košice – Čierna nad Tisou štátna hranica“ sa vyskytol v trénovacej databáze iba raz, a to projekt „ŽSR, Zavedenie ERTMS v úseku Bratislava - Žilina - Čadca - št.hr. SR/ČR“. Implementácia GSM-R navyše nie je ani projekt výlučne železničnej infraštruktúry, ale vo významnej miere obsahuje aj informačné systémy, ktorých hodnotenie už nespadá pod Oddelenie dopravných projektov ÚHP, ale pod Oddelenie projektov informatizácie a obrazy ÚHP a v tejto informačno-technologickej časti ani ÚHP nemohlo vyhodnotiť prime-

ranosť ceny z dôvodu nízkeho detailu rozpočtu[6]. Nízku presnosť a benchmarkovateľnosť tohto projektu sme preto nepovažovali za chybu nášho modelu, ale za nedostatočný detail samotného rozpočtu.

Rekonštrukcia tunela, akou je projekt „*Rekonštrukcia Bratislavského tunela č. 2, ŽST Bratislava hl. st.*“, bol unikátny prípad. Tunely sú zväčša súčasťou väčších modernizácií, pri nich ale platí existencia širokej škály vyvíjajúcich sa technológií a sú vhodným kandidátom pre žltú knihu FIDIC kvôli ťažšiemu detailnému odhadovaniu rozsahu projektu[4]. Nízky benchmarkovaný objem a vysoká relatívna výchylka kvôli nedostatkom dát bola aj v tomto prípade očakávaná.

Po zohľadnení silných stránok jednotlivých spôsobov odhadovania jednotkových cien, t. j. použitím metriky wMAPE, výsledné hodnoty nášho modelu ukazujú výrazné zlepšenie oproti pohľadu agregovane po rozpočtoch (presnosť 27,05 % pohľadom agregovane po rozpočtoch klesla na 12,60 % po zohľadnení veľkosti benchmarkovaného objemu) a porovnateľne dobre ako benchmarkovanie cenovými databázami SFDI (12,60 % našim modelom oproti 12,28 % cenovými databázami SFDI). To potvrdilo našu hypotézu o zlepšujúcej sa presnosti zvyšujúcim sa počtom dát.

V tab. 4.8 sme ešte uviedli výsledky benchmarkovania testovacej dátovej sady jednotlivými spôsobmi odhadovania jednotkových cien na vybraných typoch profesijných odborov identifikovaných v tab. 4.2 na str. 53. Náš model si viedol najlepšie zo všetkých porovnávaných spôsobov odhadovania jednotkových cien s hodnotou MAPE na vybraných typoch profesijných odborov 17,34 %, kým s ocenenými výkazmi výmer projektantmi bola MAPE 18,26 %, cenovými databázami SFDI 19,65 % a cenovou databázou CENEKON 36,79 %.

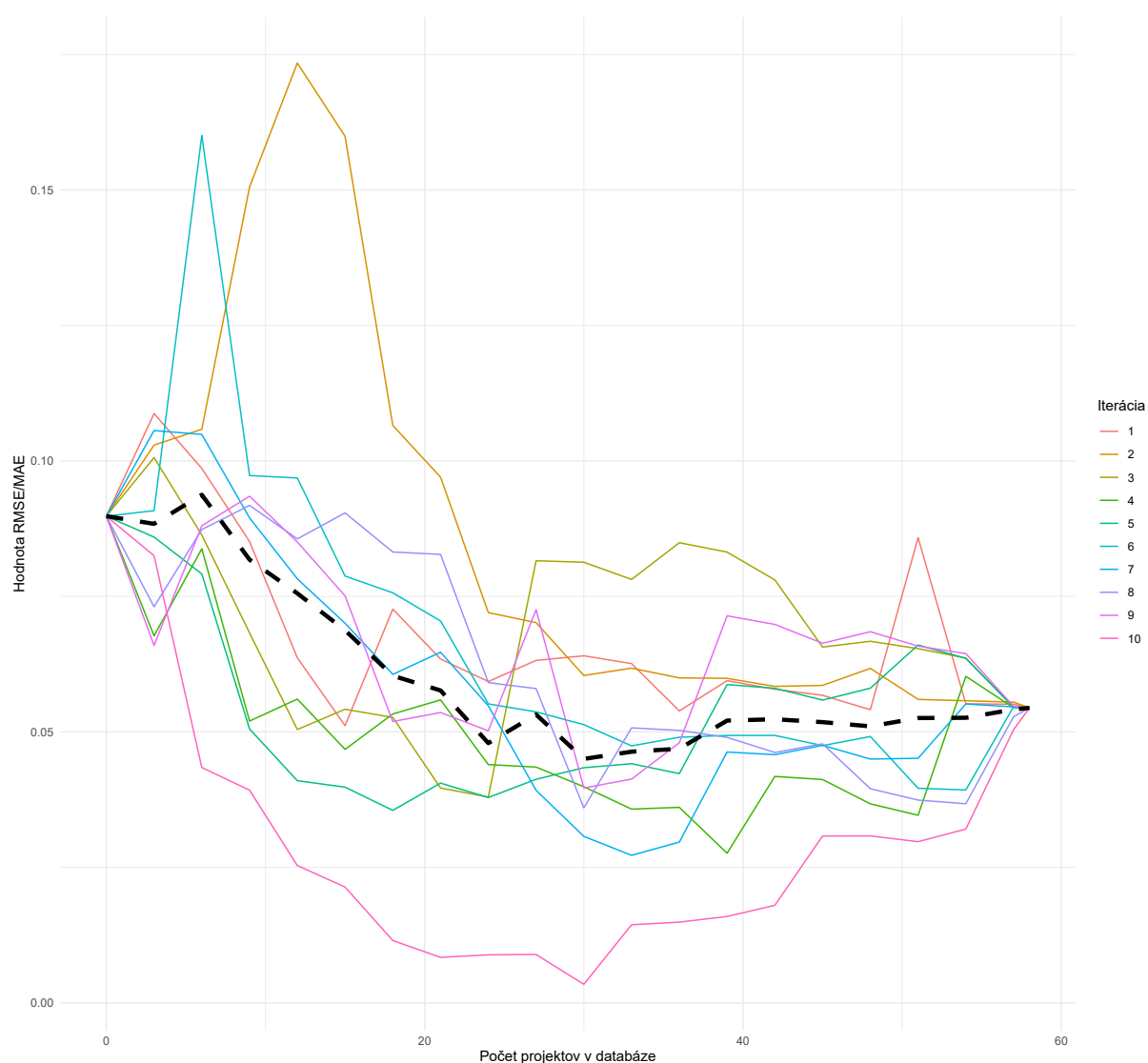
Tab. 4.8: Výsledky benchmarkovania podľa vybraných profesijných odborov

	PHZ	Model		SFDI		CENEKON	
Kód odboru	APE	Benchmarkovaný objem	APE	Benchmarkovaný objem	APE	Benchmarkovaný objem	APE
21	19,86 %	66,34 %	4,70 %	62,35 %	23,90 %	9,80 %	24,69 %
32	13,28 %	74,37 %	14,23 %	61,83 %	6,93 %	55,00 %	8,71 %
33	21,88 %	58,72 %	34,67 %	41,97 %	29,82 %	59,14 %	82,29 %
35	18,03 %	51,74 %	15,75 %	61,44 %	17,96 %	10,56 %	31,45 %

Zdroj: vlastné spracovanie

4.9 Analýza robustnosti cenovej databázy

V poslednej časti našej práce sme vykonali analýzu robustnosti našej cenovej databázy. Zaujímalo nás, ako sa menila presnosť nášho modelu popísaného v kapitole 4.7 na testovacej dátovej sade pri postupnom zväčšovaní vstupných dát pridávaním projektov. Výsledky našej analýzy sme graficky znázornili na obr. 4.4. Vykreslili sme si aj priemernú krivku 10 iterácií znázorňujúcu priemernú zmenu presnosti pri postupne pridaných projektoch. Samotný výpočtový proces analýzy robustnosti cenovej databázy v programe R trval 24 hodín a 17 minút.



Obr. 4.4: Hodnota RMSE/MAE v závislosti od počtu projektov v databáze

Zdroj: vlastné spracovanie

Z obr. 4.4 na str. 66 je viditeľný jasný klesajúci trend do počtu približne 30 projektov v cenovej databáze s následnou stabilizáciou. Dôvodom stabilizácie trendu bolo dosiahnutie dostatočnej presnosti pre projekty, akými sú obnovy a modernizácie tratí a chýbajúce dáta z projektov, ktoré sú technicky špecifické a bolo ich zatiaľ v našej cenovej databáze nedostatok. Analýza robustnosti cenovej databázy ale potvrdila, že postupným rozširovaním cenovej databázy sa presnosť odhadovania jednotkových cien zvyšuje.

Záver

V predloženej diplomovej práci sme sa venovali problematike odhadovania jednotkových cien stavebných prác a materiálov v železničných investičných projektoch. Primárnym cieľom tejto práce bolo pre potreby Oddelenia dopravných projektov ÚHP navrhnuť vlastný model odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov, ktorý by bol efektívnejší než súčasný spôsob odhadovania týchto jednotkových cien Oddelením dopravných projektov ÚHP s lepšou alebo aspoň porovnateľnou úrovňou presnosti. Pre dosiahnutie nášho primárneho cieľu sme potrebovali zrealizovať viaceré kroky, pre ktoré sme si vytýčili parciálne ciele, a to získať relevantné vstupné dáta do modelu, zaistiť rovnocennosť jednotlivých projektov vstupujúcich do modelu, zjednotiť cenovú úroveň projektov vstupujúcich do modelu, dosiahnuť skutočne relevantnú výpovednú hodnotu modelu, spresniť odhad jednotkových cien modelom, optimalizovať odhadované jednotkové ceny modelom a validovať navrhovaný model.

V prvej kapitole našej práce sme najprv rámcovo popísali problematiku odhadu nákladov železničných investičných projektov. Predstavili sme súčasný proces odhadovania nákladov investičných projektov na Slovensku počas ich projektovej prípravy. Popísali sme špecifikácie predpokladaných, ako aj reálnych rozpočtov železničných investičných projektov a problémy, aké Oddelenie dopravných projektov ÚHP pri práci s nimi musí riešiť.

V druhej kapitole sme uviedli potrebné poznatky zo štatistiky pre analýzu našich dát a tvorbu matematického aparátu nášho navrhnutého modelu. Kľúčové poznatky zo štatistiky pre našu prácu boli zaužívané metódy na identifikáciu nestability v dátach, metódy na identifikáciu vychýlenosti dát, metódy na identifikáciu outlierov v dátach, metódy na odhadovanie bodovej hodnoty, ako aj metódy na validovanie odhadu a meranie jeho presnosti.

V tretej kapitole sme sa venovali metodike analyzovaniu dát a tvorby vlastného modelu odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov. Pre potrebu relevantných vstupných dát do modelu sme vytvorili databázu jednotkových cien stavebných prác a materiálov zo zmluvných rozpočtov už realizovaných železničných investičných projektov. Pre rovnocennosť jednotlivých projektov vstupujúcich do modelu sme zostrojili vlastný mechanizmus štandardizácie váh projektov. Pre

zjednotenie cenovej úrovne projektov vstupujúcich do modelu sme navrhli vlastný systém indexácie podľa názvov objektov, ktorý bol nezávislý od nepovinných údajov v zmluvných rozpočtoch a zároveň reflektoval systém klasifikácie stavieb ŠÚ SR. Pre dosiahnutie skutočne relevantnej výpovednej hodnoty modelu sme identifikovali položky s následným očistením dát od nich, u ktorých by odhadovanie cien nebolo relevantné. Pre spresnenie odhadu jednotkových cien modelom sme identifikovali outliere s následným očistením dát od nich. Pre optimalizovanie odhadovaných jednotkových cien modelom sme matematicko-štatistickými metódami vylepšovali odhadovanie jednotkových cien vlastnou metódou optimálny percentil. Pre validáciu navrhovaného modelu sme náš navrhnutý model aplikovali na dáta simulujúce reálne podmienky v praxi a to v konfrontácii s ocenenými výkazmi výmer železničných investičných projektov od projektantov, súčasným spôsobom benchmarkovania jednotkových cien stavebných prác a materiálov železničných investičných projektov Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI, ako aj s jednotkovými cenami stavebných prác a materiálov najpoužíwanejšou komerčnou cenovou databázou na slovenskom stavebnom trhu CENEKON.

V poslednej štvrti kapitole sme zosumarizovali výsledky našej analýzy dát a výber vhodných metód pre náš model odhadovania jednotkových cien. Samotnú analýzu dát sme robili v programe R a všetky nami vytvorené skripty pre potreby našej práce sme zverejnili na internetovom hostingu GitHub [27]. Analýza systémov indexácie prostredníctvom úhrnného výberového variančného koeficientu preukázala, že náš systém indexácie podľa názvov objektov bol porovnateľne dobrý, ako systém indexácie podľa kódov odborov používaný Oddelením dopravných projektov ÚHP (tab. 4.1 na str. 50). Vzhľadom na nízky objem indexácie podľa indexu 124 Budovy pre dopravu a telekomunikáciu v systéme názvov objektov a súčasne veľmi nesignifikantný rozdiel pri znížení počtu indexov z 5 na 3 odporúčame ÚHP zvážiť vylúčenie aj tohto indexu. Analýzou vychýlenosti dát sme zistili pozitívnu vychýlenosť našich dát (obr. 4.2 na str. 55), čo nám pomohlo pri výbere vhodnej metódy identifikujúcej outliere, ktorou sa ukázala byť metóda Adjusted Boxplot (kapitola 4.4 na str. 56-57). Analýza existencie objemnej zľavy nepreukázala, že by sme mali v dátach objemnú zľavu, iba naznačila určitý sklon k nej (tab. 4.5 na str. 58). Nakoniec sme krížovou validáciou a inšpirujúc sa metódou zlatého rezu nakalibrovali vlastnú metódu odhadovania jednotkových cien optimálny percentil a metriku merania presnosti MAPE

sme preukázali, že je presnejšia, ako iné bežne používané metódy aritmetický priemer či medián (tab. 4.6 na str. 63). Výsledkom našej analýzy bol návrh konkrétneho modelu odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov popísaného v kapitole 4.7 na str. 62.

Na záver sme validovali navrhnutý model na dátach simulujúcich reálne podmienky v praxi, kde sme preukázali výrazne zrýchlenie práce až o 99,97 %, vyšší benchmarkovateľný objem o 12,66 % a vyššiu presnosť odhadovania jednotkových cien o 49,37 % ako súčasný spôsob benchmarkovania železničných investičných projektov Oddelením dopravných projektov ÚHP pomocou cenových databáz SFDI a aj ako komerčná cenová databáza CENEKON (tab. 4.7 na str. 64). Stochastickou simuláciou sme nakoniec ešte demonštrovali zlepšujúcu sa presnosť odhadovania jednotkových cien pri rozširovaní cenovej databázy (obr. 4.4 na str. 66).

Prínosom práce bolo okrem samotného navrhnutia modelu odhadovania jednotkových cien stavebných prác a materiálov železničných investičných projektov aj jej okamžité praktické využitie na Oddelení dopravných projektov ÚHP, ako aj jej potenciál využitia ďalšími štátnym inštitúciami Slovenskej republiky pracujúcich so železničnými projektami či nové poznatky do oblasti vyvíjania cenových databáz a benchmarkovania jednotkových cien s možnou adaptáciou na širšie spektrum investičných projektov, akými sú napr. projekty cestnej infraštruktúry či pozemných stavieb. Prínosom pre autora bolo oboznámenie sa s plánovaním a realizáciou investičných projektov na Slovensku, ako aj rozšírenie poznatkov v oblasti štatistiky a strojového učenia s ich priamou aplikáciou na reálne dáta.

Možným rozšírením našej práce by bolo navrhnutý model s vytvorenou cenovou databázou skombinovať s inými databázami jednotkových cien stavebných prác a materiálov, čím by sa jednoznačne zvýšil benchmarkovateľný objem, ako aj potenciálne zlepšila presnosť odhadovania týchto jednotkových cien. Práca tiež ponúka aj možnosti ďalšieho výskumu v oblasti analýzy existencie objemnej zľavy a možnej transformácie navrhnutého statického modelu na dynamický, ktorý by pri benchmarkovaní vždy individuálne priradil jednotkovú cenu na základe konkrétneho množstva danej položky v predloženom predpokladanom rozpočte. Potenciál je aj v zefektívnení práce s cenovou databázou možným exportovaním cenovej databázy do systému pre manažment databázy (angl. *database management system*).

Zoznam použitej literatúry

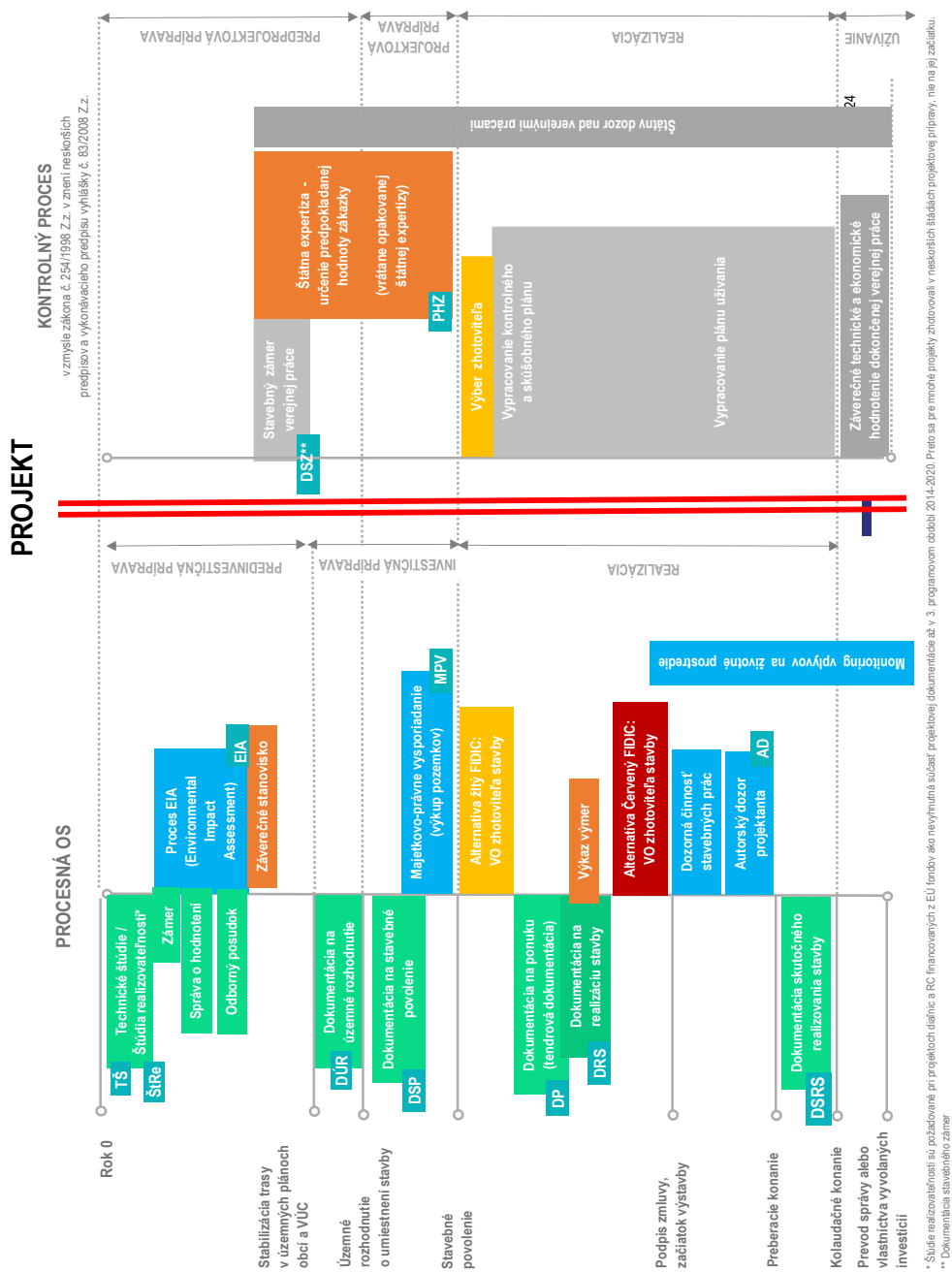
- [1] Arlot S., Celisse A.: *A survey of cross-validation procedures for model selection*, Statistics Surveys 4 (2010), 40-79, dostupné na internete (1.4.2023): <https://doi.org/10.1214/09-SS054>
- [2] ASB: *Použitie metódy Naprojektuj a postav 2.*, anketa, 2012, dostupné na internete (16.4.2023): <https://www.asb.sk/stavebnictvo/inzinierske-stavby/mosty/anketa-pouzitie-metody-naprojektuj-a-postav-2>
- [3] *Centrálny register zmlúv*, internetová stránka, dostupné na internete (8.1.2023): <https://www.crz.gov.sk/>
- [4] Deloitte: *Metodika pro zadání veřejné zakázky formou „DESIGN - BUILD“ pro dopravní stavby v ČR*, závěrečná správa, SFDI Ministerstvo dopravy České republiky, 2015, dostupné na internete (18.4.2023): https://www.sfdi.cz/soubory/obrazky-clanky/metodiky/2015_metodika_db_zaverecna_zprava.pdf
- [5] Ellingerová H.: *Triednik stavebných prác a jeho aplikácia v praxi*, Nehnutelnosti a bývanie 2 (2007), 20-25, dostupné na internete (16.4.2023): https://www.stuba.sk/new/docs/stu/ustavy/ustav_manazmentu/NAB2007_2/4clanok.pdf
- [6] Farkaš R., Rako M.: *Hodnota za peniaze projektu Implementácia GSM-R do siete ŽSR, úsek Varín – Košice – Čierna nad Tisou*, hodnotenie, ÚHP MF SR, Bratislava, 2022, dostupné na internete (29.4.2023): https://www.mfsr.sk/files/archiv/20/Hodnotenie-ZSR-GSM-R-Varin-KE-CnT_20220518.pdf
- [7] Harman R.: *Analýza zhlukov a klasifikácia dát*, prednášky, FMFI UK, Bratislava, 2022
- [8] Hastie T., Tibshirani R., Friedman J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer New York, New York, 2009
- [9] Hyndman R. J., Athanasopoulos, G.: *Forecasting: principles and practice*, OTexts, Melbourne, 2018

- [10] Hyndman R., Koehler A.: *Another look at measures of forecast accuracy*, International Journal of Forecasting 22 (2006), 679-688, dostupné na internete (13.4.2023): <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- [11] Iglewicz B., Hoaglin D.: *How to detect and handle outliers*, ASQC Quality Press, Milwaukee, 1993
- [12] Katina S., Králik M., Hupková A.: *Aplikovaná štatistická interferencia I*, Masarykovská univerzita, Brno, 2015, dostupné na internete (30.4.2023): <https://munispace.muni.cz/library/catalog/view/538/1682/290-1/#preview>
- [13] Kišš Š. et al.: *Revízia výdavkov na dopravu*, záverečná správa, ÚHP a IFP MF SR, Bratislava, 2016, dostupné na internete (13.4.2023): https://www.mfsr.sk/files/archiv/uhp/3370/76/Finalna_sprava_revizia_DOPRAVA.pdf
- [14] Kohavi R.: *A study of cross-validation and bootstrap for accuracy estimation and model selection*, International Joint Conference on Artificial Intelligence 2 (1994), 1137-1143, dostupné na internete (8.4.2023): <https://dl.acm.org/doi/10.5555/1643031.1643047>
- [15] Kondorová L.: *Osobná komunikácia*, Oddelenie posudzovania cien ŽSR, Košice, 2022
- [16] *Metodický pokyn Ministerstva dopravy a výstavby Slovenskej republiky č. 19/2022, ktorým sa stanovuje mechanizmus úpravy ceny v dôsledku zmien nákladov pri projektoch opravy a údržby, výstavby, modernizácie a rekonštrukcie inžinierskych stavieb a budov*
- [17] MF SR: *Hodnota za peniaze*, internetová stránka, dostupné na internete (14.1.2023): <https://www.mfsr.sk/sk/financie/hodnota-za-peniaze/>
- [18] Moore D. S., Notz W. I., Flinger M. A.: *The Basic Practice of Statistics*, Purdue University, W. H. Freeman and Company, West Lafayette, 2013
- [19] *Nariadenie vlády Slovenskej republiky č. 174/2019 Z. z. o podmienkach vypracovania štúdie uskutočniteľnosti investície a štúdie uskutočniteľnosti koncesie*

- [20] *Nariadenie vlády Slovenskej republiky č. 330/2018 Z. z., ktorým sa ustanovuje výška sadzieb poplatkov za uloženie odpadov a podrobnosti súvisiace s prerozdeľovaním príjmov z poplatkov za uloženie odpadov*
- [21] Ott R. L., Longnecker M.: *An Introduction to Statistical Methods and Data Analysis*, Texas A&M University, Cengage Learning, College Station, 2010
- [22] Rako M.: *Indexácia železničných investičných projektov podľa ŽSR (aktualizácia)*, interný dokument, Oddelenie dopravných projektov ÚHP MF SR, Bratislava, 2023
- [23] Rako M.: *Osobná komunikácia*, Oddelenie dopravných projektov ÚHP MF SR, Bratislava, 2022-2023
- [24] Rako M., Starek T.: *Rešerš verejných obstarávaní investičných projektov železničnej infraštruktúry za roky 2018 až 2022*, interný dokument, Oddelenie dopravných projektov ÚHP MF SR, Bratislava, 2023
- [25] Seo S.: *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*, diplomová práca, University of Pittsburgh, Pittsburgh, 2006, dostupné na internete (10.1.2023): <http://d-scholarship.pitt.edu/7948/1/Seo.pdf>
- [26] Somorčík J.: *Počítačová štatistika*, prednášky, FMFI UK, Bratislava, 2020
- [27] Starek T.: *Zverejnené použité algoritmy v programe R pre potreby diplomovej práce*, GitHub, 2023, dostupné na internete (30.4.2023): <https://github.com/Handynew1999/DP>
- [28] ŠÚ SR: *Indexy cien stavebných prác podľa Klasifikácie stavieb (2015=100) - štvrťročne*, tabuľka v databáze DATAcube, ŠÚ SR, dostupné na internete (31.3.2023): <https://datacube.statistics.sk/>
- [29] Taboga M.: *Training, validation and test samples*, článok, StatLect, 2021, dostupné na internete (30.3.2023): <https://www.statlect.com/machine-learning/training-validation-and-test-samples>
- [30] Trnovská M.: *Metódy voľnej optimalizácie*, prednášky, FMFI UK, Bratislava, 2020

- [31] Turzák P., Starek T.: *Určovanie nákladov cestných projektov podľa historických dát*, interný dokument, Oddelenie dopravných projektov ÚHP MF SR, Bratislava, 2021
- [32] *Uznesenie vlády Slovenskej republiky č. 649/2020 k návrhu rozpočtu verejnej správy na roky 2021 až 2023*
- [33] *Úrad pre verejné obstarávanie*, internetová stránka, dostupné na internete (8.1.2023): <https://www.uvo.gov.sk>
- [34] Vanderviere E., Huber M.: *An adjusted boxplot for skewed distributions*, COMPS-TAT'2004 Symposium (2004), 1933-1940, dostupné na internete (10.1.2023): <https://wis.kuleuven.be/stat/robust/papers/2004/boxplotCOMPSTAT04.pdf>
- [35] *Vyhláška Štátnej komisie pre techniku č. 107/1966 o dokumentácii stavieb*
- [36] Witten H. I., Frank E., Hall A. M.: *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*, Morgan Kaufmann, Boston, 2011
- [37] *Zákon č. 254/1998 Z. z. o verejných prácach*
- [38] *Zákon č. 523/2004 Z. z. o rozpočtových pravidlách verejnej správy a o zmene a doplnení niektorých zákonov*

A Prvá príloha



Obr. A.1: Plánovanie a príprava investičných projektov na Slovensku

Zdroj: [13]

B Druhá príloha

Zoznam projektov v tréningovej databáze:

- „ČNT, KR výhybiek č. 145š-147š, 150š-152š, 154š -159š, 161š, 162š, 14 výhybiek“
- „Elektrifikácia trate Haniska pri Košiciach – Moldava nad Bodvou“
- „Hul – Úľany, sanácia zemného telesa v km 34,850-35,050“
- „Komplexná rekonštrukcia EPZ Zvolen“
- „Komplexná rekonštrukcia mosta v žkm 98,530 trate Devínska Nová Ves – Štúrovo“
- „KRŽZ kolají č. 1 v ŽST Snina a ŽST Stakčín“
- „Maťovce ŠRT št. hr. UZ – Maťovce ŠRT k. č. 1š“
- „Modernizácia žel. trate Nové Mesto nad Váhom - Púchov, úsek: Beluša – Púchov“
- „Modernizácia železničnej trate Žilina – Košice, úsek trate Liptovský Mikuláš – Poprad Tatry (mimo), 1.etapa (Poprad - Lučivná)“
- „Náhrada telemechanizačného zariadenia duplexného TZD-751 v diaľkovo ovládaných dopravných OR Zvolen“
- „Nákl. Nemčanka, komplexná rekonštrukcia na dopravňu DOD“
- „NŽST Ulanka, rekonšt. výhybiek č. 5, 8, 9, 10 a sanácia žel. spodku a diaľk. ovládanie zab. zar. v úseku B. Bystrica – Ulanka; NŽST Ulanka – staničné zab. zar.“
- „Odstránenie nedostatočnej šuntovej citlivosti KO na PZZ-OR Trnava“
- „Odstránenie nedostatočnej šuntovej citlivosti KO na PZZ, OR Zvolen“
- „Pial – Beša, sanácia zemného telesa km 17,8-19,4“
- „Pomodernizačné úpravy Dobudovanie protihlukovej steny v ŽST Trenčín – vetva CH“
- „Pribeňík- Velký Horeš, KRŽZ, kol. č. 2“

- „Púchov – Lúky pod Makytou, KR trakčného vedenia a zabezpečovacieho zariadenia; ŽST Lúky pod Makytou, KR trakčného vedenia; Lúky pod Makytou – št. hr. ČR, KR trakčného vedenia“
- „Rekonštrukcia mosta v km 51,368 trate Devínska Nová Ves - Štúrovo“
- „Rekonštrukcia NŽST Martin“
- „Spínacia stanica Veľký Horeš“
- „TNS Galanta, komplexná rekonštrukcia rozvodne 110kV“
- „Traťový úsek Dvory nad Žitavou – Nové Zámky, odstránenie mimoriadneho stavu zemného telesa k. č.2 km 150,050 – 150,250“
- „Trenčín – Trenčianska Turná – Mníchova Lehota KRŽŽ kol. č. 1“
- „Výh. Budkovce ŠRT kolaj číslo 3š, 5š“
- „Výh. Slivník ŠRT rekonštrukcia kolaje 1š, 2aš, 4š, 6š“
- „Výh. Slivník ŠRT – Výh. Slančík ŠRT, KRŽŽ km 59,190-62,690“
- „Výh. ŠTRKOVEC, KR kol. č. 1, km 6,343 – 7,162“
- „Výhybňa Milochov - železničný zvršok“
- „Výmena KO 25 Hz za počítače osí - ŽST Bajč, ŽST Nové Zámky, medzistaničný úsek Bajč – Nové Zámky“
- „Zriadenie zabezpečeného priechodu pre peších trate Košice – Žilina žkm 131,253“
- „Zvyšovanie bezpečnosti na železničných priecestiach traťový úsek Čadca – Makov, PZZ v km 9,076“
- „Zvyšovanie stability násypového telesa trate ŠRT Maťovce – ŠRT Haniska pri Košiciach v km 41,515 – 41,565 a v km 41,640 – 41,701“
- „ŽSR Čierna nT, obnova kolají č. 824 a 914š mat. užitým“

- „ŽSR, Dostavba zriaďovacej stanice Žilina-Teplice a nadväzujúcej železničnej infraštruktúry v uzle Žilina“
- „ŽSR, Modernizácia trate Púchov - Žilina, pre rýchlosť do 160 km/hod., I. etapa (úsek Púchov - Považská Teplá)“
- „ŽSR, Modernizácia trate Púchov - Žilina, pre rýchlosť do 160 km/hod., II. etapa - (úsek Považská Teplá /mimo/ - Žilina /mimo/)“
- „ŽSR, rekonštrukcia výhybiek č. 1 a 2 v ŽST Velké Uherce“
- „ŽSR, rekonštrukcia výhybiek č. 5 a 6 v ŽST Chynorany“
- „ŽSR, rekonštrukcia výhybiek č.21, a 25 v ŽST Lužianky“
- „ŽSR, Terminál integrovanej osobnej prepravy Trebišov“
- „ŽSR, Zavedenie ERTMS v úseku Bratislava - Žilina - Čadca - št.hr. SR/ČR“
- „ŽSR, ŽST Prievidza – ŽST Handlová, rekonštrukcia k. č. 1, km 15,336-16,570“
- „ŽST Bratislava hl. st., prístupy pre imobilných cestujúcich na nástupištia“
- „ŽST Dolný Štál, predĺženie staničných kolají“
- „ŽST Jablonica KRŽZ, k.č.2,4,4a“
- „ŽST Jesenské – rekonštrukcia výhybiek č. 1, 3XA, 5, 7, 7XA, 8, 11 a rekonštrukcia nástupíšť“
- „ŽST Košice, obnova výhybiek č. 298XA, 301XA, 301, 305, 306, 307, 308, 310, 311, 315, 309, 314“
- „ŽST Košice – rekonštrukcia nástupíšť č. 1, 2, 3“
- „ŽST Kralovany – ŽST Lubochňa, rekonštrukcia k. č. 1,2“
- „ŽST Maťovce, predĺženie kolaje a KR výhybiek s úpravou ŽZ, zab. zar. s TV“
- „ŽST Maťovce ŠR – VÝH Vojany ŠR, rekonštrukcia kol. č. 1 v dĺžke 3,2 km“
- „ŽST Nižná Myšľa – obnova výhybiek č. 10, 11, 12, 13, 14, 15 a stred DKS“

- „ŽST Podbrezová, rekonštrukcia výhybiek“
- „ŽST Rimavská Seč, rekonštrukcia výhybiek“
- „ŽST Rimavská Sobota, rekonštrukcia výhybiek“
- „ŽST Trebišov – výhybňa Hrinište, KR k.č.1, materiálom novým, km 20,737-24,707;
3,970 km, sanácia žel. spodku v úseku 20,737-21,880“
- „ŽST Vrútky - Varín, koľ. č. 1 a 2 KRŽZ km 317,457 - 328,730“

C Tretia príloha

Zoznam projektov v testovacej databáze:

- „*Boleráz - Smolenice, KRŽZ koľ. č. 1*“
- „*Implementácia GSM-R do siete ŽSR, úsek Varín – Košice – Čierna nad Tisou štátna hranica*“
- „*Rekonštrukcia Bratislavského tunela č. 2, ŽST Bratislava hl. st.*“
- „*ŽSR, Elektrifikácia a súvisiace rekonštrukcie na úseku trate Bánovce nad Ondavou – Humenné*“
- „*ŽSR, Modernizácia koridoru, štátna hranica ČR/SR - Čadca - Krásno nad Kysucou (mimo), železničná trať, 3.etapa*“
- „*ŽSR, ŽST Dvory nad Žitavou, rekonštrukcia výhybiek č. 14, 15, 16, 17, 18, 19, 20*“
- „*ŽST Barca, obnova výhybiek č. 1-20*“
- „*ŽST Nové Zámky – ŽST Palárikovo, rekonštrukcia koľaje č. 1 a č. 2*“
- „*ŽST Trnovec nad Váhom – ŽST Tvrdošovce, rekonštrukcia koľaje č. 1 a koľaje č. 2*“

D Štvrtá príloha

Tab. D.1: Výsledky benchmarku pre model – po projektoch

Projekt	Zmluvná cena	PHZ		Model		
		Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
Rekonštrukcia kolaje č. 1 Boleráz – Smolenice	7 023 885,86 €	6 327 297,72 €	9,92 %	6 143 107,86 €	64,04 %	12,54 %
Implementácia GSM-R Varín – Košice – Čierna nad Tisou št. hr.	2 756 845,49 €	3 063 059,54 €	11,11 %	5 383 920,99 €	7,04 %	95,29 %
Rekonštrukcia Bratislavského tunela č. 2	1 445 446,86 €	1 639 573,53 €	13,43 %	2 598 237,44 €	23,71 %	79,7 5%
Elektrifikácia trate Bánovce nad Ondavou – Humenné	117 827 443,42 €	103 449 390,08 €	12,20 %	102 225 057,68 €	64,54 %	13,24 %
Modernizácia trate Svrčinovec – št. hr.	50 331 166,18 €	51 694 898,55 €	2,71 %	57 923 006,61 €	68,38 %	15,08 %
Rekonštrukcia výhybiek č. 14 až 20 v stanici Dvory nad Žitavou	4 467 605,38 €	3 480 860,12 €	22,09 %	4 500 239,77 €	41,41 %	0,73 %
Obnova výhybiek č. 1 až 20 v stanici Barca	4 678 627,96 €	4 057 419,73 €	13,28 %	4 159 968,74 €	36,65 %	11,09 %
Rekonštrukcia kolají č. 1 a 2 Nové Zámky – Palárikovo	22 677 972,46 €	18 894 398,13 €	16,68 %	21 949 164,83 €	64,75 %	3,21 %
Rekonštrukcia kolají č. 1 a 2 Trnovec nad Váhom – Tvrdošovce	24 937 593,01 €	22 334 777,94 €	10,44 %	21 815 380,55 €	70,31 %	12,52 %
SPOLU	236 146 586,61 €	214 941 675,33 €	8,98 %	226 698 084,45 €	58,11 %	4,00 %

Zdroj: vlastné spracovanie

Tab. D.2: Výsledky benchmarku pre SFDI – po projektoch

Projekt	Zmluvná cena	PHZ		SFDI		
		Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
Rekonštrukcia kolaje č. 1 Boleráz – Smolenice	5 694 213,05 €	5 390 154,58 €	5,34 %	6 782 446,52 €	51,91 %	19,11 %
Implementácia GSM-R Varín – Košice – Čierna nad Tisou št. hr.	12 985 643,73 €	13 569 315,40 €	4,49 %	13 732 270,93 €	33,18 %	5,75 %
Rekonštrukcia Bratislavského tunela č. 2	3 781 653,78 €	2 204 067,76 €	41,72 %	4 012 290,27 €	62,03 %	6,10 %
Elektrifikácia trate Bánovce nad Ondavou – Humenné	89 444 429,88 €	76 843 214,29 €	14,09 %	97 798 449,35 €	49,00 %	9,34 %
Modernizácia trate Svrčinovec – št. hr.	37 780 871,56 €	33 276 547,71 €	11,92 %	45 864 522,08 €	51,33 %	21,40 %
Rekonštrukcia výhybiek č. 14 až 20 v stanici Dvory nad Žitavou	4 937 953,54 €	4 039 795,93 €	18,19 %	4 341 948,23 €	45,78 %	12,07 %
Obnova výhybiek č. 1 až 20 v stanici Barca	8 341 752,66 €	7 484 717,24 €	10,27 %	5 850 586,54 €	65,35 %	29,86 %
Rekonštrukcia kolají č. 1 a 2 Nové Zámky – Palárikovo	22 786 528,20 €	17 864 689,51 €	21,60 %	21 413 915,61 €	65,06 %	6,02 %
Rekonštrukcia kolají č. 1 a 2 Trnovec nad Váhom – Tvrdošovce	23 879 941,87 €	21 048 349,64 €	11,86 %	26 405 863,09 €	67,33 %	10,58 %
SPOLU	209 632 988,28 €	181 720 852,08 €	13,31 %	226 202 292,60 €	51,58 %	7,90 %

Zdroj: vlastné spracovanie

Tab. D.3: Výsledky benchmarku pre CENEKON – po projektoch

Projekt	Zmluvná cena	PHZ		CENEKON		
		Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
Rekonštrukcia kolaje č. 1 Boleráz – Smolenice	4 486 099,25 €	3 975 756,40 €	11,38 %	4 297 563,42 €	40,90 %	4,20 %
Implementácia GSM-R Varín – Košice – Čierna nad Tisou št. hr.	3 257 956,84 €	2 172 669,02 €	33,31 %	2 217 001,67 €	8,33 %	31,95 %
Rekonštrukcia Bratislavského tunela č. 2	1 946 603,62 €	2 554 160,07 €	31,21 %	2 577 802,30 €	31,93 %	32,43 %
Elektrifikácia trate Bánovce nad Ondavou – Humenné	55 004 436,67 €	54 745 662,72 €	0,47 %	47 214 756,36 €	30,13 %	14,16 %
Modernizácia trate Svrčinovec – št. hr.	40 411 043,47 €	48 649 613,99 €	20,39 %	63 930 493,43 €	54,90 %	58,20 %
Rekonštrukcia výhybiek č. 14 až 20 v stanici Dvory nad Žitavou	1 970 497,12 €	1 705 097,72 €	13,47 %	1 833 289,29 €	18,27 %	6,96 %
Obnova výhybiek č. 1 až 20 v stanici Barca	7 723 335,21 €	6 821 023,99 €	11,68 %	6 502 315,25 €	60,51 %	15,81 %
Rekonštrukcia kolají č. 1 a 2 Nové Zámky – Palárikovo	12 911 938,78 €	11 400 970,38 €	11,70 %	12 641 969,18 €	36,87 %	2,09 %
Rekonštrukcia kolají č. 1 a 2 Trnovec nad Váhom – Tvrdošovce	14 890 891,64 €	14 177 832,16 €	4,79 %	15 226 759,48 €	41,98 %	2,26 %
SPOLU	142 602 802,61 €	146 202 786,44 €	2,52 %	156 441 950,38 €	35,09 %	9,70 %

Zdroj: vlastné spracovanie

Tab. D.4: Výsledky benchmarku pre model – po profesijných odboroch

		PHZ		Model		
Kód odboru	Zmluvná cena	Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
21	32 565 483,28 €	25 586 508,42 €	21,43 %	31 034 310,57 €	66,34 %	4,70 %
22	10 375 978,91 €	9 661 909,34 €	6,88 %	14 903 414,86 €	40,51 %	43,63 %
23	14 274,95 €	11 519,04 €	19,31 %	10 889,80 €	2,76 %	23,71 %
24	295 724,14 €	212 365,54 €	28,19 %	253 059,66 €	2,68 %	14,43 %
25	198 739,55 €	461 721,80 €	132,33 %	134 020,10 €	1,60 %	32,56 %
26	4 345,50 €	4 038,63 €	7,06 %	3 450,52 €	14,76 %	20,60 %
27	535 686,71 €	599 732,13 €	11,96 %	560 107,23 €	50,01 %	4,56 %
29	705 386,36 €	580 476,92 €	17,71 %	686 377,43 €	57,95 %	2,69 %
31	1 774 707,71 €	1 868 048,75 €	5,26 %	1 605 762,56 €	64,50 %	9,52 %
32	110 678 124,14 €	95 224 968,62 €	13,96 %	94 924 443,72 €	74,37 %	14,23 %
33	30 842 671,17 €	38 992 648,22 €	26,42 %	41 536 230,35 €	58,72 %	34,67 %
34	5 661 410,97 €	6 063 148,72 €	7,10 %	4 979 211,13 €	27,28 %	12,05 %
35	38 028 970,89 €	31 084 392,20 €	18,26 %	32 038 774,19 €	51,74 %	15,75 %
36	852 945,01 €	652 229,94 €	23,53 %	708 466,88 €	78,48 %	16,94 %
37	1 184 209,18 €	1 339 286,73 €	13,10 %	1 174 496,28 €	45,59 %	0,82 %
38	2 312 802,60 €	2 430 470,05 €	5,09 %	1 977 659,13 €	71,96 %	14,49 %
39	115 125,55 €	168 210,28 €	46,11 %	167 410,05 €	73,63 %	45,42 %

Zdroj: vlastné spracovanie

Tab. D.5: Výsledky benchmarku pre SFDI – po profesijných odboroch

		PHZ		SFDI		
Kód odboru	Zmluvná cena	Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
21	30 609 035,16 €	24 817 839,06 €	18,92%	37 923 450,64 €	62,35 %	23,90 %
22	4 110 871,07 €	3 831 519,63 €	6,80 %	6 003 863,25 €	16,05 %	46,05 %
23	182 076,76 €	153 531,25 €	15,68 %	287 228,55 €	35,16 %	57,75 %
24	1 690 949,50 €	1 064 196,60 €	37,07%	1 334 470,92 €	15,32 %	21,08 %
25	3 581 760,14 €	3 893 604,40 €	8,71 %	4 594 421,66 €	28,86 %	28,27 %
26	0 €	0 €	NaN	0 €	0,00 %	NaN
27	0 €	0 €	NaN	0 €	0,00 %	NaN
29	0 €	0 €	NaN	0 €	0,00 %	NaN
31	2 181 604,66 €	2 176 133,96 €	0,25 %	3 359 284,77 €	79,29 %	53,98 %
32	92 018 491,10 €	77 469 933,09 €	15,81 %	98 396 399,00 €	61,83 %	6,93 %
33	22 043 153,36 €	23 570 948,53 €	6,93 %	28 616 692,69 €	41,97 %	29,82 %
34	7 979 903,13 €	7 659 156,89 €	4,02 %	8 540 636,03 €	38,46 %	7,03 %
35	45 158 739,83 €	37 013 771,08 €	18,04 %	37 050 234,24 €	61,44 %	17,96 %
36	12 113,07 €	9 688,79 €	20,01 %	9 702,52 €	1,11 %	19,90 %
37	33 155,37 €	27 623,78 €	16,68 %	44 617,37 €	1,28 %	34,57 %
38	31 135,13 €	32 905,00 €	5,68 %	41 290,96 €	0,97 %	32,62 %
39	0 €	0 €	NaN	0 €	0,00 %	NaN

Zdroj: vlastné spracovanie

Tab. D.6: Výsledky benchmarku pre CENEKON – po profesijných odboroch

Kód odboru	Zmluvná cena	PHZ		CENEKON		
		Benchmark	APE	Benchmark	Benchmarkovaný objem	APE
21	4 811 053,15 €	3 980 959,51 €	17,25%	3 623 234,76 €	9,80 %	24,69 %
22	2 953 757,39 €	2 545 521,82 €	13,82 %	2 270 001,63 €	11,53 %	23,15 %
23	17 470,30 €	13 903,14 €	20,42 %	12 406,92 €	3,37 %	28,98 %
24	207 094,53 €	150 031,71 €	27,55 %	143 677,04 €	1,88 %	30,62 %
25	56 233,65 €	53 489,84 €	4,88 %	58 313,41 €	0,45 %	3,70 %
26	20 015,09 €	16 593,11 €	17,10 %	14 852,82 €	68,00 %	25,79 %
27	204 257,11 €	256 345,47 €	25,50 %	179 208,99 €	19,07 %	12,26 %
29	18 125,47 €	14 978,97 €	17,36 %	13 586,44 €	1,49 %	25,04 %
31	2 277 608,45 €	2 124 063,01 €	6,74 %	2 207 945,20 €	82,78 %	3,06 %
32	81 862 047,26 €	73 853 120,71 €	9,78 %	74 728 745,31 €	55,00 %	8,71 %
33	31 064 226,60 €	45 573 871,37 €	46,71 %	56 628 461,15 €	59,14 %	82,29 %
34	7 404 867,96 €	6 899 564,13 €	6,82 %	6 944 900,10 €	35,69 %	6,21 %
35	7 761 614,91 €	6 657 980,94 €	14,22 %	5 320 376,48 €	10,56 %	31,45 %
36	442 990,92 €	282 501,39 €	36,23 %	277 234,35 €	40,76 %	37,42 %
37	971 074,01 €	1 094 044,26 €	12,66 %	1 229 962,89 €	37,38 %	26,66 %
38	2 408 931,41 €	2 510 077,30 €	4,20 %	2 544 403,59 €	74,95 %	5,62 %
39	121 434,40 €	175 739,77 €	44,72 %	244 639,30 €	77,66 %	101,46 %

Zdroj: vlastné spracovanie