

ŽILINSKÁ UNIVERZITA V ŽILINE

---

**Autoreferát**  
DIZERTAČNEJ PRÁCE

---

Žilina, máj 2020

Ing. Milan Straka

Žilinská univerzita v Žiline  
Fakulta riadenia a informatiky

Ing. Milan Straka

Autoreferát dizertačnej práce

**Analýza rozsiahlych dát  
v dopravných a energetických aplikáciách**

---

Názov dizertačnej práce

na získanie akademického titulu „**philosophiae doctor**“ (PhD.)  
v študijnom programe doktorandského štúdia  
**inteligentné informačné systémy**

v študijnom odbore  
**informatika**

Žilina, máj 2020

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre matematických metód a operačnej analýzy, Fakulte riadenia a informatiky Žilinskej univerzity v Žiline

**Predkladateľ**      Ing. Milan Straka  
Žilinská univerzita v Žiline  
Fakulta riadenia a informatiky  
Katedra matematických metód a operačnej analýzy

**Školiteľ**            prof. Ing. Luboš Buzna, PhD.  
Žilinská univerzita v Žiline  
Fakulta riadenia a informatiky  
Katedra matematických metód a operačnej analýzy

**Oponenti**            doc. Ing. Roman Jarina, PhD.  
Žilinská univerzita v Žiline  
Fakulta elektrotechniky a informačných technológií  
Katedra multimédií a informačno-komunikačných technológií

prof. Ing. Vladimír Gazda, PhD.  
Technická univerzita v Košiciach  
Ekonomická fakulta  
Katedra financií

Autoreferát bol rozoslaný dňa: .....

Obhajoba dizertačnej práce sa koná dňa **19. augusta 2020** o **10:30** hod. pred komisiou pre obhajobu dizertačnej práce schválenu odborovou komisiou v študijnom odbore **informatika**, v študijnom programe **inteligentné informačné systémy**, vymenovanou dekanom Fakulty riadenia a informatiky Žilinskej univerzity v Žiline dňa

.....

**prof. RNDr. Jaroslav Janáček, CSc.**  
predseda pracovnej skupiny odborovej komisie  
študijného programu **inteligentné informačné systémy**  
v študijnom odbore **informatika**  
Fakulta riadenia a informatiky  
Žilinská Univerzita  
Univerzitná 8215/1  
010 26 Žilina

# Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Súčasný stav v problematike</b>	<b>2</b>
2.1	Základné pojmy . . . . .	2
2.2	Používané metódy . . . . .	3
2.2.1	Regresné metódy . . . . .	4
2.2.2	Klasifikačné metódy so schopnosťou regresie . . . . .	4
2.2.3	Metódy časových radov . . . . .	6
2.2.4	Zhlukovacie metódy . . . . .	6
2.3	Použité dáta . . . . .	6
<b>3</b>	<b>Ciele dizertačnej práce</b>	<b>7</b>
<b>4</b>	<b>Vlastný prístup k riešeniu problematiky</b>	<b>9</b>
4.1	Analýza segmentov nabíjacích staníc a zákazníkov . . . . .	9
4.1.1	Použité dáta a spracovanie dát . . . . .	9
4.1.2	Výsledky zhlukovania . . . . .	11
4.2	Predikovanie spotreby nabíjacích staníc . . . . .	12
4.2.1	Použité dáta a ich spracovanie . . . . .	13
4.2.2	Modelovanie a predikcia spotreby elektrickej energie . . . . .	13
4.3	Extrakcia prediktorov z GIS dát a testovanie metód na výber premenných . . . . .	14
4.4	Modelovanie popularity . . . . .	16
4.5	Modelovanie energie . . . . .	20
4.5.1	Modelovanie rozdelenia spotreby energie . . . . .	21
4.5.2	Vysvetľovanie spotreby energie ostatnými indikátormi nabíjacích miest . . . . .	21
4.5.3	Vysvetlenie spotreby energie na nabíjacích miestach pomocou GIS dát . . . . .	22
4.5.4	Vplyv stratégie umiestňovania na spotrebu energie . . . . .	25
	<b>Publikácie</b>	<b>31</b>
	<b>Zoznam použitej literatúry</b>	<b>32</b>

# 1 Úvod

Žijeme v dobe, kde dáta tvoria jeden z najcennejších zdrojov a sú často označované aj ako ropa 21. storočia [6]. Dokážeme ich zbierať takmer všade a spracovávať vo veľkých množstvách. Spracovaniu a analýze dát sa postupne venuje čoraz viac ľudí a celý svet sa, či už priamo alebo nepriamo, v každodennom živote stretáva s aplikáciami umelej inteligencie a strojového učenia.

O elektromobilite sa v súčasnosti hovorí ako o riešení, ktoré môže prieniesť veľký potenciál vo viacerých smeroch, akými sú, napríklad, nižšie emisie, vyššia efektivita elektrických motorov oproti spalovacím alebo nezávislosť od fosílnych palív. V tejto práci elektromobilitu vnímame aj ako moderný príklad prepojenia energetických a dopravných systémov.

Proces nasadzovania elektromobility je spojený s rôznymi výzvami a komplikáciami. Napríklad, cena elektrických vozidiel je výrazne vyššia ako cena bežných vozidiel. Takisto z dôvodu ich nabíjania je potrebné najskôr vybudovať samostatnú infraštruktúru a aby elektromobilita výraznejšie zlepšovala životné prostredie, mala by elektrická energia pochádzať z obnoviteľných zdrojov. Ak bude rozhodovanie v elektromobilite vhodne podporené dátami, ktoré začínajú byť dostupné, môžeme predpokladať, že budú ušetrené nezanedbateľné finančné prostriedky. Celý proces pri budovaní dátami poháňaných rozhodnutí v prostredí elektromobility sa dá rozdeliť do viacerých fáz. Prvou fázou je vytvorenie plánu a stanovenie cieľov. Druhou fázou je zber, spracovanie, analýza a vyhodnotenie týchto dát. Posledná fáza je vytvorenie rozhodnutí na základe získaných výsledkov, alebo vykonanie, napríklad, ďalších simulačných či optimalizačných štúdií a následné vyvodenie záverov a rozhodnutí. V súlade s týmito trendmi si táto práca kladie za cieľ navrhnúť postup, akým bude možné podporiť rozhodovanie v tejto oblasti a to konkrétne využitím prostriedkov analýzy dát.

## 2 Súčasný stav v problematike

### 2.1 Základné pojmy

#### Pojmy týkajúce sa analýzy dát

Analýza dát je multidisciplinárna oblasť spájajúca niekoľko oblastí, ako napr. matematika, informatika a kontext skúmanej oblasti, u nás elektromobilita. Pre prácu s dátami definujeme nevyhnutné základné pojmy.

*Vysvetľujúca premenná*  $X$ , nazývaná tiež vstupná, je náhodná premenná, často viacrozmerná. *Vysvetľovaná premenná*  $Y$ , nazývaná tiež výstupná, je náhodná premenná, ktorú vysvetľujeme pomocou vstupnej premennej alebo premenných. Vzťah medzi výstupnou a vstupnou premennou popisujeme funkciou  $Y = f(X) + \epsilon$ , kde  $f$  je funkcia vstupnej premennej a  $\epsilon$  je náhodná chyba [16, s. 16]. Za jednu z najdôležitejších úloh dátového modelovania môžeme považovať nájdenie vhodnej reprezentácie funkcie  $f$ . *Dátová matica*, ako realizácia vstupnej premennej  $\mathbf{X}$  rozmeru  $n \times p$  je tvorená  $i = 1, \dots, n$  pozorovaniami  $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$  (riadky matice  $\mathbf{X}$ ) a *prediktormi*  $x_j = \{x_{1j}, \dots, x_{nj}\}$  pre  $j = 1, \dots, p$  (stĺpce matice  $\mathbf{X}$ ). Vektor realizácií výstupnej premennej, označíme  $y$ , a budeme ho nazývať vektor výstupu. Dáta predpokladáme v určitej štandardnej forme, ktorú nazývame *dataset*.

## Elektromobilita

Terminológia pre elektromobilitu je na Slovensku definovaná v Zákone č. 251/2012 Z. z. Zákon o energetike a o zmene a doplnení niektorých zákonov, §2b [26], my ju doplníme pomocou definícií pojmov elektromobility [25]. Pod pojmom *elektrické vozidlo* (EV) budeme rozumieť vozidlo, ktoré má elektromotor schopný samostatne poháňať vozidlo. Osobu, ktorá jazdí na EV, či už firemné alebo vlastné, budeme jednotne označovať ako používateľa EV. Pod *nabíjajúcim bodom* rozumieme rozhranie, ktoré v určitom čase umožňuje nabíjanie jedného EV. *Konektor* je nabíjacia zásuvka, na ktorú sa môže napojiť práve jedno vozidlo. Nabíjací bod môže mať jeden alebo viac konektorov. *Nabíjacia stanica* je jeden alebo viac nabíjajúcich bodov, integrovaných do jedného zariadenia. *Nabíjajúcim miestom* budeme podľa [25] rozumieť miesto, na ktorom sa nachádza jedna alebo viac nabíjajúcich staníc, zvyčajne umiestnených vedľa seba. *Nabíjacou transakciou*, zjednodušene len transakciou, budeme rozumieť nabíjaciu udalosť začínajúcu pripojením a končiacu odpojením vozidla. Pojem *smart charging* je v [13] definovaný ako optimalizácia času, rýchlosti a smeru nabíjacej udalosti.

## 2.2 Používané metódy

Z dátovej analýzy využívame najmä tejto fáze využívame najmä metódy strojového učenia. Strojové učenie je subdoménou umelej inteligencie, ktorá využíva učenie, ako napríklad zlepšovanie modelov na základe vlastných skúseností bez explicitného naprogramovania [24]. Strojové učenie možno rozdeliť podľa prítomnosti vektora výstupu – učiteľa na dve základné ka-

tegorie: učenie s učiteľom a učenie bez učiteľa.

Pri učení s učiteľom súčasne využívame vopred definované prediktory a aj vektor výstupu. Cieľom je modelovať vzťah medzi vektorom výstupu a prediktormi čo najpresnejšie tak, aby tento vzťah mohol byť použitý na predpovedanie budúcich pozorovaní (predikcia) alebo lepšie porozumenie súvislostí medzi vektorom vstupu a prediktormi (inferencia) [16]. Typy úloh strojového učenia môžeme deliť podľa typu výstupnej premennej na úlohy klasifikačné, ktorých výstupné premenné sú diskkrétne a úlohy regresné, ktorých výstupné premenné sú spojité.

Pri učení bez učiteľa nemáme pre prediktory vopred definovaný vektor výstupu, takže nás nemá kto „učiť“. Snažíme sa preto hľadať vzťahy medzi prediktormi alebo medzi pozorovaniami a nájsť tak napríklad podobné skupiny medzi nimi. Príkladom takéhoto učenia je zhluková analýza.

### 2.2.1 Regresné metódy

Regresné metódy predpokladajú spojitú výstupnú premennú. Pri regresných metódach v tejto práci využívame najmä schopnosť inferencie.

#### Tolerantné metódy

Tolerantné metódy redukovujú koeficienty lineárnej regresie a poskytujú tak riešenie s väčším vychýlením no menším rozptylom. Optimalizačný problém pre takéto metódy možno všeobecne zapísať matematickým modelom ako minimalizáciu sumy štvorcov rezíduí (angl. residual sum of squares - RSS) a pokutovej funkcie

$$\underset{\beta}{\operatorname{argmin}} \operatorname{RSS} + \lambda g(\cdot), \quad (1)$$

pričom  $\lambda \geq 0$  je hyperparameter, ktorý kontroluje mieru redukcie,  $g(\cdot)$  je pokutová funkcia, najčastejšie kombinácia alebo modifikácia  $l$ -normiem. Čím väčšiu hodnotu nadobúda  $\lambda$ , tým viac sú koeficienty redukované. Lasso [16, s. 219] využíva ako pokutovú funkciu  $g(\cdot)$   $l_1$  normu, t.j.  $g(\cdot) = \lambda \sum_{j=1}^p |\beta_j|$ , kde  $\beta_j$  je regresný koeficient.

### 2.2.2 Klasifikačné metódy so schopnosťou regresie

#### Lineárne klasifikačné metódy

Vybrané klasifikačné metódy predikujú pravdepodobnosť a správajú sa v tomto zmysle podobne ako regresné metódy [16]. Lineárne klasifikačné metódy modelujú výstupnú premennú pomocou lineárnych vzťahov so

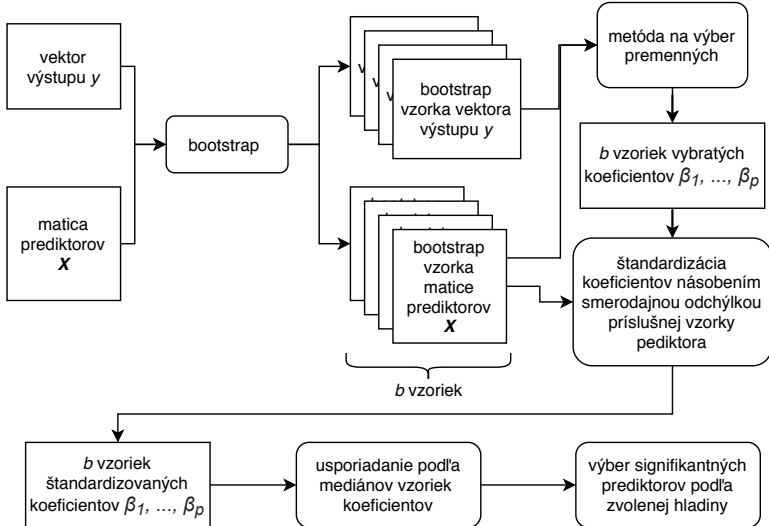
vstupmi. Pre logistickú regresiu s LR  $l_1$  (LR- $l_1$ ) regularizáciou je optimalizačný problém definovaný ako

$$\min_{\beta_0, \beta} -\frac{1}{n} \sum_{i=1}^n \left\{ y_i (\beta_0 + \beta^T \mathbf{x}_i) - \log(1 + e^{\beta_0 + \beta^T \mathbf{x}_i}) \right\} + \lambda \sum_{j=1}^p |\beta_j|, \quad (2)$$

kde  $\beta_0$  je odstup a  $\beta^T$  je vektor koeficientov. Pre klasifikáciu pravdepodobnostnej hodnoty na binárnu využívame prahový parameter  $\theta$ .

## Štatistická inferencia po výbere premenných

Klasické  $p$ -hodnoty, získané pre odhady pomocou OLS, nemusia po dátovo riadenom výbere premenných byť ďalej platné, keďže je model vybratý stochastickým procesom, napr. tolerantnou metódou, čo klasická štatistická teória negarantuje [1]. Pre lasso a LR- $l_1$  je odhad  $p$ -hodnôt vďaka adaptívnej procedúre komplikovaný. Aby nedošlo k prezentácii nespoľahlivých výsledkov, a aj na základe odporúčaní v podobnej štúdií [14], tak pre získanie informácie o signifikantnosti koeficientov vykonávame štatistickú inferenciu pomocou metódy bootstrap [39, s. 142], ktorej postup ilustrujeme na obrázku 1.



Obr. 1: Diagram štatistickej inferencie pomocou metódy bootstrap.



## Stromové metódy

V práci sme využili dve metódy založené na rozhodovacích stromoch. Prvou je random forest (RF) [11, s. 587], ktorá využíva baggingu (bootstrap aggregation), kombináciu bootstrapu a agregácie [16, s. 319]. Princíp metódy spočíva v tom, že využíva viacero stromov – les, trénovaných na vzorkách dát bootstrpom, a v každom rozhodovacom uzle využíva náhodnú podmnožinu prediktorov. Druhou je gradient boosted regression trees (GBRT) [11, s. 379], využívajúca boosting, iteratívne vylepšujúc rezíduá z predošlých jednoduchých stromov.

### 2.2.3 Metódy časových radov

Na základe literatúry [5, 20] aj experimentov sme využívali sezónne ARIMA modely s exogénnymi premennými (SARIMAX) na modelovanie časových radov, schopné vhodne zachytiť sezónnosť. Ako exogénnu premennú sme použili aj Fourierove rady [15]. Chybu meriame pomocou priemernej absolútnej percentuálnej chyby (angl. mean absolute percentage error) [15].

### 2.2.4 Zhlukovacie metódy

Pre zhlukovú analýzu využívame tri metódy. Prvou je  $k$ -means [10, s. 451], ktorá iteratívne priraduje pozorovania k centráam zhlukov. Ďalšou je aglomeračné hierarchické zhlukovanie [10, s. 457], využívajúce spôsob zdola nahor, zlučujúce iteratívne pozorovania na základe vzdialenosti do zhlukov a zhluky do väčších zhlukov, pokiaľ nevytvoria všetky pozorovania jeden spoločný zhluk. Poslednou je DBSCAN [10, s. 471], ktorý spája pozorovania na základe hustoty, čím sa značne líši od zvyšných algoritmov.

## 2.3 Použité dáta

Spoločnosť ElaadNL nám poskytla dva datasety pochádzajúce z nabíjacej infraštruktúry v Holandsku, spoločne nazývané EVnetNL. Dataset *Transactions* obsahuje dáta z nabíjacích transakcií, ako napr. RFID karta používateľa, čas pripojenia, čas nabíjania, maximálny výkon transakcie, nabitá energia a identifikátor nabíjacej stanice. Dataset *Meterreadings* obsahuje 15 minútové odčítania z merača pri nabíjaní a dáta, ako napr. nabitá energia, maximálny výkon a identifikátor transakcie.

### GIS dáta

Tabuľka 1 poskytuje prehľad použitých dát geografických informačných systémov (GIS). Uvažujeme 4 základné reprezentácie dát [17, s. 55]: ras-

trové, vektorové bodové, vektorové lomené čiary a vektorové polygóny.

Dataset	Stručný popis	reprezentácia dát
Populačné jadrá	Detailné populačné dáta na úrovni obcí a miest [36].	polygóny
Susedstvá	Populačné dáta agregované do susedstiev [35].	polygóny
Využitie územia	25 kategórií využitia územia vo vysokom rozlíšení [34].	polygóny
Energia	Agregovaná spotreba energií na úrovni susedstiev [8].	polygóny
Životná úroveň	Všeobecný index popisujúci kvalitu života na úrovni susedstiev [23].	polygóny
Dopravné toky	Objem dopravy na jednotlivých cestných úsekoch [40].	lomené čiary
LandScan	Ambientná populácia v rastrovom formáte. [27]	raster
OpenStreetMap	OpenStreetMap (OSM) body záujmu. [29]	body
Nabíjacie stanice 2015	Dostupné umiestnenia nabíjacích staníc z roku 2015. [28, 30]	body

**Tab. 1:** Prehľad zozbieraných GIS dát spolu s ich typom

### 3 Ciele dizertačnej práce

Dáta môžu pomôcť pri identifikácii správnych rozhodnutí avšak na to ich potrebujeme vhodne spracovať a vyhodnotiť. Na tento účel slúžia rôzne nástroje a metódy, ktorých je v dnešnej dobe veľké množstvo. Tieto nástroje a metódy je potrebné vedieť správne používať, prípadne rozšíriť existujúce metódy alebo vytvoriť vhodné kombinácie metód pre dosiahnutie čo najpriaznivejších výsledkov. Na základe rokovaní s expertmi zo spoločnosti ElaadNL [7], holandského znalostného a inovačného centra v oblasti nabíjacej infraštruktúry a spoločnosti Greenway [9], ktorá sa venuje budovaniu a prevádzke nabíjacích staníc na Slovensku a okolitých krajinách, ako aj prehľadu dostupnej literatúry a prác v oblasti elektromobility, sme formulovali tri ciele pre dizertačnú prácu. Hlavným cieľom je vytvoriť metodológiu pre analýzu dát a dátové modelovanie v prostredí elektromobility, ktorá bude mať potenciál zväčšiť množinu dostupných nástrojov pre podporu rozho-

dovania v oblasti budovania a prevádzky nabíjacej infraštruktúry pre EV, pričom sa budeme sústreďovať na tri nižšie uvedené čiastkové ciele.

### **Ciel 1: Vytvorenie segmentov nabíjacích staníc a zákazníkov**

Na začiatku má o elektromobilitu záujem iba určitá skupina ľudí – tzv. skorí inovátori a postupne vznikajú nové kategórie zákazníkov, ktoré je potrebné identifikovať, sledovať a reagovať ponukou služieb na ich dopyt. Podobne nabíjacie stanice sú využívané rôznym spôsobom, nočné nabíjanie v blízkosti domova, denné nabíjanie v blízkosti práce a pod. A teda segmenty zákazníkov, príp. nabíjacích staníc, sú charakterizované tým, že vykazujú iný vzorec nabíjania elektrických vozidiel.

Našou úlohou je na základe vizuálnych analýz dát a zhlukovacích algoritmov identifikovať zhluky a porovnať ich s literatúrou. Výstupom má byť lepšie pochopenie nabíjacieho správania a takisto nájdené segmenty môžu byť využité na stratifikáciu dát v ďalších analýzach, ako napr. predikcia energie alebo predikcia využívania určitej skupiny staníc.

### **Ciel 2: Predikovanie spotreby elektrickej energie na nabíjacích staniciach**

Z hľadiska prevádzkovateľa staníc, ale aj prevádzkovateľa elektrickej siete, je potrebné predikovať budúci dopyt po elektrickej energii, či už z krátkodobého alebo aj dlhodobého hľadiska. Tu sa jedná najmä o vytvorenie podpory pre operatívne rozhodnutia, ako sú napríklad nákup elektrickej energie pre budúce obdobie alebo aj predpoveď spotreby na ďalší deň pre potreby smart chargingu. Otestujeme možnosti vytvoriť krátkodobé a dlhodobé predpovede spotreby elektrickej energie na nabíjacích staniciach a uvážime rôznu mieru agregácie staníc.

Spotrebu chceme predpovedať pomocou odhadových metód určených pre časové rady a metód strojového učenia. Modely sa pokúsime vylepšiť pomocou rôznych externých premenných ako napríklad počasie v okolí stanice. Okrem priestorovej agregácie vyskúšame aj agregáciu staníc na základe nabíjacích vlastností, čo by mohlo zlepšiť presnosť predpovedí. Výsledkom môžu byť okrem vhodnej metodiky na predpovedanie agregovanej spotreby staníc aj informácie o variabilite spotreby a miery do akej vieme túto spotrebu predpovedať.

### **Ciel 3: Identifikácia ukazovateľov vhodného umiestnenia nabíjacích staníc**

Jedným zo základných prvkov v elektromobilite sú nabíjacie stanice. Ich umiestňovanie je súčasťou strategických rozhodnutí, ktoré sú spojené s

vyššími finančnými nákladmi a dlhodobou platnosťou týchto rozhodnutí. Z prehľadov literatúry vieme, že stanice sa často umiestňujú na základe empiricky podložených, alebo čiastočne empiricky podložených rozhodnutí, pričom je nedostatok literatúry venujúcej sa charakteristikám vhodného umiestnenia takýchto staníc.

Miesta vhodné pre umiestnenia nabíjajúcich staníc, a najmä ich charakteristiky, chceme určovať na základe existujúcich dát o nabíjaní vozidiel, spolu so socio-ekonomickými ukazovateľmi a rôznymi inými druhmi dát. Na odhad vplyvu plánujeme využiť inferenčnú schopnosť lineárnej regresie a jej modifikácií. Z týchto modifikácií využijeme najmä tolerantné a selekčné metódy pretože sa dá očakávať, že prediktory odvodené z dát budú často podobné a závislé, s čím sa tieto metódy vedú čiastočne vysporiadať.

Očakávame, že na základe analýz budeme vedieť identifikovať faktory, ktoré potenciálne vplývajú na nabíjanie na stanicach. Modely využívajúce tieto faktory ako prediktory, by mali byť schopné, určiť vhodné umiestnenia staníc a teda výsledky bude možné použiť pre rozšírenie lokálnych analýz.

## 4 Vlastný prístup k riešeniu problematiky

Táto kapitola je venovaná opisu navrhnutých riešení a vlastných prínosov, v súlade so stanovenými cieľmi uvedenými v predošlej časti.

### 4.1 Analýza segmentov nabíjajúcich staníc a zákazníkov

Zhlukovanie v dostupných štúdiách skúmajúcich nabíjajúcu infraštruktúru slúžilo najmä na analýzu segmentov zákazníkov a popis nabíjacieho správania. My sa pozrieme na analýzu segmentov nabíjajúcich staníc, ktorá je v literatúre pokrytá v podstatne menšom rozsahu ako segmentácia zákazníkov [31, 42]. Poznanie nabíjajúcich staníc s podobným správaním napomáha k lepšiemu porozumeniu využívania nabíjacej infraštruktúry. Okrem samotnej segmentácie porovnávame aj dva prístupy na získanie zhlukov (segmentov) staníc z dát vo forme transakcií, líšiacich sa najmä v agregáciách.

#### 4.1.1 Použité dáta a spracovanie dát

Na základe vytvoreného prehľadu literatúry sme identifikovali tri hlavné triedy indikátorov na charakterizovanie výkonnosti nabíjajúcich staníc a

transakcií: návštevnosť stanice, využitie stanice a časové vzory používania stanice. Pre každú triedu sme vybrali minimálnu množinu indikátorov. Pri transakciách uvažujeme pre návštevnosť  $N_{tran}$  – počet transakcií na stanici, kde prebehla transakcia. Z pohľadu využívania  $r_{tran}$  – relatívny čas nabíjania, t.j. pomer medzi časom nabíjania a časom pripojenia. A z pohľadu časového vzoru používania  $TS_{tran}$  – čas začiatku transakcie v hodinách a  $TE_{tran}$  – čas konca transakcie v hodinách. Pri nabíjaciach stanicach pre návštevnosť  $N_{stat}$  – počet transakcií na stanici. Z pohľadu využívania  $r_{stat}$  – relatívny čas nabíjania, t.j. pomer medzi časom nabíjania a časom pripojenia vypočítaný zo všetkých transakcií, ktoré na stanici prebehli. Pre časový vzor používania  $TS_{stat}$  – priemerný čas začiatku všetkých transakcií v hodinách a  $TE_{stat}$  – priemerný čas konca všetkých transakcií v hodinách, ktoré na nabíjacej stanici prebehli.

### **Použité dáta a spracovanie dát**

Na výpočet indikátorov sme použili iba dáta medzi 1. januárom 2015 a 31. decembrom 2015, kedy bol počet transakcií pomerne stály a obdobie 1 roka nie je tak ovplyvňované sezónnosťou, ako by to mohlo byť napr. pri pätnástich mesiacoch. Aby sme získali reprezentatívne dáta, neuvažovali sme stanice, ktoré mali menej ako 30 transakcií. Pri analýze zvolených indikátorov sme identifikovali multimodalitu, ktorá na jednej strane spôsobuje straty informácie spôsobené agregáciou transakcií do staníc, najmä pri časových indikátoroch, na druhej strane ukazuje, že by zhlukovacie algoritmy mohli vrátiť rozlíšiteľné segmenty staníc.

Na zhlukovanie využívame metódy  $k$ -means, aglomeratívne hierarchické zhlukovanie a DBSCAN. Z dôvodu, že metódy vyžadujú vstupy v rovnakých jednotkách, škálujeme vstupné hodnoty na rozsah  $(0, 1)$ . Kvôli rôznemu počtu transakcií na stanicach vznikla otázka, ako správne reprezentovať stanice pomocou dát z transakcií. Preto navrhujeme a porovnávame dva nasledovné prístupy.

*Agregácia prvá - zhlukovanie druhé* - tento prístup najskôr sumarizuje indikátory transakcií pre stanice, spočítaním priemerov indikátorov transakcií, ktoré prebehli na stanici. V ďalšom kroku aplikuje zhlukovací algoritmus na takto sumarizované dáta staníc. Tento prístup skrátene označíme ako *agregačný* prístup, pretože oproti nasledovnému prístupu dáta agreguje.

*Zhlukovanie prvé - kategorizácia druhá* - prístup najskôr zhlukovacími algoritmi nájdeme zhľuky transakcií a potom priradí stanice k zhľukom, v ktorých majú najvyšší počet transakcií. Toto priradenie je asociované s veľkou neistotou, ak je pre stanicu podobný počet transakcií v dvoch

alebo viacerých zhlukoch. Tento prístup skrátene označíme ako *priradovací* prístup, pretože oproti predošlému prístupu priraduje stanice k zhlukom.

#### 4.1.2 Výsledky zhlukovania

Pre agregáčny prístup DBSCAN algoritmus našiel dva malé zhluky a jeden veľký zhluk, pričom nechal veľa staníc nepriradených žiadnemu zo zhlukov, čo napovedá, že metóda nie je vhodná na takýto typ dát a tak sme ju vylúčili z ďalších analýz. Pre priradovací prístup je potrebné vypočítavať maticu vzdialeností pre obe metódy okrem  $k$ -means čo je pri takmer 300 000 pozorovaniach výpočtovo náročné, preto sme ich pri tomto prístupe neaplikovali.

V metóde  $k$ -means sme nastavili parameter  $k = 4$  na základe metódy lakta [10, s. 486] pre oba prístupy a v aglomeratívnom hierarchickom zhlukovaní bola použitá miera kompletného prepojenia [10, s. 462].

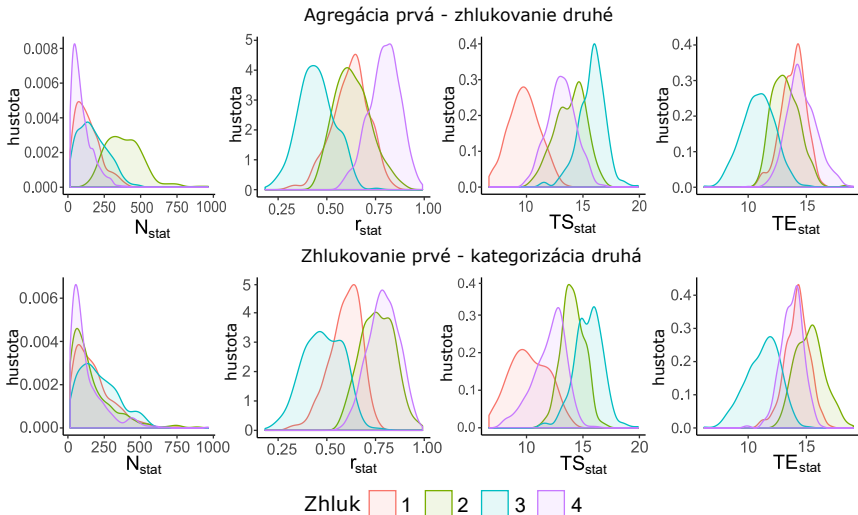
Pre agregáčny prístup boli výsledky  $k$ -means a aglomeratívneho hierarchického veľmi podobné, preto aj z dôvodu lepšieho porovnania s priradovacím prístupom porovnáваме iba výsledky  $k$ -means zhlukovania.

#### Porovnanie prístupov

Najskôr sme porovnali priradenie staníc k zhlukom medzi prístupmi, kde sa väčšina prvkov zhodovala. Funkcie hustôt zhlukov zobrazujeme na obrázku 2. Na základe hodnôt atribútov sme interpretovali zhluky nasledovne. Zhluk č. 1 reprezentuje nabíjacie stanice, na ktorých prevláda nabíjanie pri zamestnaní, kde nabíjanie začína ráno a končí večer, zhluk č. 2 sú prevažne krátke poobedné nabíjania s vysokým využitím času pripojenia na nabíjanie a pri agregáčnom prístupe aj vysokým počtom transakcií na stanici. Tento zhluk môže reprezentovať napríklad nabíjanie pri supermarketoch alebo obchodných centrách. Zhluk č. 3 je pravdepodobne nabíjanie v obytných oblastiach, kde sa EV nabíjajú cez noc, pretože sa nabíjanie začína večer, končí ráno a má nízky pomer nabíjania k času pripojenia. V zhluku č. 4 nabíjanie začína okolo obeda a končí poobede, pričom pre agregáčny prístup má nižší pomer nabíjania k času pripojenia oproti priradovaciemu prístupu, ktorý má tento pomer vysoký.

Predpokladáme, že pozorované rozdiely v priradení staníc k zhlukom sú spôsobené multimodalitou rozdelenia pomeru času nabíjania a pripojenia. Zhluky získané metódou *agregácia prvá zhlukovanie druhé* sú ľahšie interpretovateľné aj vďaka hodnotám  $N_{tran}$ , ktoré sa viacej variujú oproti druhej metóde.

Na záver podkapitoly môžeme konštatovať, že porovnanie oboch prí-



**Obr. 2:** Funkcie hustoty indikátorov nabíjajúcich staníc priradených do rovnakého zhluku metódou k-means. Vrchný rad prislúcha k zhlukom nabíjajúcich staníc obdržaných agregáčnym prístupom, spodný k zhlukom priradovacieho prístupu.

stupov ukázalo, že obe metódy identifikovali podobné skupiny staníc. Z dvoch porovnaných metodologických prístupov, považujeme za lepší prístup *agregovanie prvé zhlukovanie druhé*, na základe jednoduchšej výpočtovej zložitosti dát a lepšej interpretovateľnosti výsledkov.

## 4.2 Predikovanie spotreby nabíjajúcich staníc

Táto podkapitola sa venuje predikovaniu časopriestorovo agregovaného dopytu nabíjajúcich staníc po elektrickej energii pomocou časových radov.

Pre prevádzkovateľov siete nabíjajúcich staníc ale aj pre dodávateľov energie, je výhodné mať čo najlepšie informácie o energii, ktorá bude spotrebovaná nabíjacou infraštruktúrou. Vďaka vopred známemu profilu a objemu elektrickej energie môžu prevádzkovatelia energiu vopred nakúpiť výhodnejšie, ale aj vylepšiť technológie smart chargingu. V [2] sa snažili krátkodobu predpovedať obsadenosť a spotrebu energie jednotlivých staníc, pričom nedosiahli uspokojivé výsledky a odporučili stanice priestorovo agregovať. Predpovede pre jednotlivé stanice sú obtiažne najmä preto, že je potrebné predpovedať typ pripojeného vozidla a obsadenosť stanice. Ak je daná stanica obsadená, používateľ EV pravdepodobne použije stanicu v blízkom okolí. Takáto zmena sa na priestorovo agregovanej spotrebe nie-

Procedúra	Trénovacia množina	Predikcia
statická	$y_1, \dots, y_i$	$y_{i+1}, \dots, y_k$
fixný počiatok	$y_1, \dots, y_{j-1}$	$y_j, \dots, y_k$
posúvajúci sa počiatok	$y_{j-i}, \dots, y_{j-1}$	$y_j, \dots, y_k$

**Tab. 2:** Trénovacie množiny a predikčné pre jednotlivé procedúry, kde  $y$  je časový rad,  $i$  je koniec prvej trénovacej množiny (čas prvej predpovede mínus 1),  $j$  je začiatok predpovede,  $k$  je koniec predikcie.

koľkých staníc nemusí prejavit.

#### 4.2.1 Použité dáta a ich spracovanie

Pre priestorovú agregáciu staníc sme použili COROP regióny, deliace Holandsko na 40 častí. Z dát sme extrahovali časové rady s dennou frekvenciou. Výrazne vyššiu ako aj stabilnejšiu spotrebu mal v porovnaní s inými COROP regiónmi región Utrecht, v ktorom má EVnetNL dataset najviac nabíjajúcich staníc (169). Pre trénovanie predikčných modelov uvažujeme obdobie medzi 3. marcom 2014 a 1. marcom 2015 (364 dní) a pre testovanie nasledovných 280 dní. Sezónnosť v dátach sme identifikovali pomocou autokorelačných grafov.

#### Externé dáta

Pre vylepšenie modelov sme uvažovali nasledovné prediktory. Binárny prediktor pre víkendy a sviatky, keďže bola spotreba v priemere o 30 % nižšia pre tieto dni. Pre viacnásobnú sezónnosť sme pridali počet super používateľov, ktorých bolo 300, 8 % z celkového počtu, no spotrebovali približne 50 % celkovej energie. Pre lepšie zachytenie sezónnosti sme pridali aj Fourierove rady.

#### 4.2.2 Modelovanie a predikcia spotreby elektrickej energie

Spotrebovanú energiu nabíjajúcich stanicami sme sa rozhodli modelovať pomocou SARIMAX, RF a GBRT modelov, chybu merať pomocou MAPE a predpovedať na 7, 14 a 28 dní. Keďže autor v [20] považuje takéto dáta za nestacionárne a odporúča pretrénovávať modely, z tohto dôvodu sme sa rozhodli vytvoriť nasledovné trénovacie procedúry, zobrazené v tabuľke 2, zohľadňujúce rôznu dĺžku trénovacích dát, ako aj predikcií.

Statická procedúra ako jediná nemení trénovaciu množinu. Pre lepšie zachytenie rozdielov medzi pracovnými dňami a víkendmi, sme pridali model *SARIMAX striedajúci*, kde sa striedali dva SARIMAX modely, prvý



MAPE modelu (%)	Dĺžka predpovede		
	7 dní	14 dní	28 dní
SARIMAX jednoduchý	<b>12.02</b>	12.35	12.97
SARIMAX striedavý	12.18	<b>12.15</b>	<b>12.60</b>
RF	12.58	12.55	13.28
GBRT	12.68	13.22	13.45
PM	16.40	16.86	17.98

**Tab. 3:** Výsledky pre jednotlivé modely a dĺžky predpovedí v percentách MAPE. Najlepšie výsledky pre jednotlivé dĺžky predpovedí sú zvýraznené tučným písmom.

pre pracovné dni, druhý pre víkendy, pričom druhý (víkendový) využíva ako exogénnu premennú priemer predikcie prvého.

Najlepšie výsledky podľa MAPE dosiahla pre SARIMAX a RF procedúra s fixným počiatkom a pre GBRT procedúra s posúvajúcim sa počiatkom, pričom sme ďalej použili na tréning metód tieto procedúry.

Výsledky predpovedí uvádzame pre kombináciu metód a tréningových procedúr v tabuľke 3. PM je perzistenčný model, čo je nulový model využívajúci sezónnosť, predikujúci periódy cieľového časového horizontu replikovaním poslednej pozorovanej periódy.

Pre 7 dňovú dĺžku predpovedí dosahuje najlepšie výsledky jednoduchý SARIMAX model s hodnotu MAPE o 1.3 menšou v porovnaní s druhým najlepším modelom a s chybou o 26.7 % menšou v porovnaní s PM. Pre 14 a 28 dňovú predpoveď dosahuje najlepšie výsledky striedavý SARIMAX model a to o 28 % a 30 % lepšie v porovnaní s PM modelom.

Na základe týchto výsledkov môžeme usúdiť, že na zvolenej časovo-priestorovej úrovni vieme s relatívne dobrou presnosťou predpovedať spotrebovanú energiu. Dosiahnuté výsledky sú približne o 30 % lepšie ako poskytuje sezónny model PM.

### 4.3 Extrakcia prediktorov z GIS dát a testovanie metód na výber premenných

Nasledujúce dve podkapitoly sa venujú analýze indikátorov vhodného umiestnenia nabíjacej infraštruktúry pomocou regresných úloh. V tejto kapitole uvedieme opis metodológie, spoločnej pre obe podkapitoly. Najskôr sa budeme zaoberať tvorbou prediktorov z GIS dát a následne predstavíme výsledky porovnania metód určených na výber premenných.

## Nabíjacie miesta

Ak sú stanice príliš blízko seba, napr. stanice v podzemných parkoviskách, nedokážeme ich poriadne rozlíšiť v priestorovej analýze a tak sme vytvorili z nich nabíjacie miesta tak, že sme iteratívne spájali všetky nabíjacie stanice v okruhu 50 metrov do jedného nabíjacieho miesta.

## Extrakcia atribútov z GIS dát

Pre extrakciu prediktorov z GIS dát sme zvolili kruhové zóny [38] (angl. round buffer). Identifikovali sme podľa [4, s. 12] niekoľko typov atribútov: atribút vyjadrujúci počet (napr. populácia), priemerný atribút (napr. priemerná mzda obyvateľov), percentuálny atribút (napr. percento žien v populácii), kategorický nominálny atribút (napr. využitie územia: obytná zóna) a poradový atribút (hustota adries, vyjadrená číslami od 1 po 5). Pre výpočty sme spoločným názvom označili ako *relatívne atribúty* priemerné, percentuálne a kategorické poradové atribúty. Pre lepšiu odhad sme pre relatívne atribúty použili *súvisiace atribúty*, ktoré vylepšujú presnosť odhadu. Vzťahy pre odhad skalárnej hodnoty z okolia nabíjacieho miesta pomocou kruhovej zóny boli založené na vzťahu pre aritmetický priemer. Tieto výpočty sme aplikovali na polygónové a rastrové dáta.

Pre každý nabíjacie miesto sme identifikovali najbližší cestný segment a použili jeho tok ako prediktor. Okrem toho sme pre každú kruhovou zónu vypočítali hustotu dopravy a ciest podľa [19].

Z bodových datasetov OpenStreetMap a Nabíjacie stanice 2015 sme extrahovali prediktory ako vzdialenosť k najbližšiemu bodu a počet bodov v kruhovej zóne.

## Testovanie metód na výber premenných

Predspracované dáta z okolia nabíjacej infraštruktúry vykazujú vysokú mieru multikolinearity, tak sme preverili, ako sa s ňou metódy na výber premenných vysporiadali. Porovnávali sme metódu najmenších štvorcov (angl. ordinary least squares - OLS), metódy krokovej selekcie: výberovú regresiu hrubou silou [16, s. 205], doprednú [16, s. 207] a spätnú regresiu [16, s. 209] a tolerantné metódy: hrebeňovú regresiu [16, s. 215], lasso [16, s. 219], elastic net [11, s. 661] a PACS [33]. Vytvorili sme umelé dáta simulujúce kolinearitu a multikolinearitu v rôznych stupňoch. Na porovnanie metód sme využili Manhattanskú vzdialenosť koeficientov [10, s. 72].

Celkovo dosiahli najlepšie výsledky metódy PACS a lasso, pričom metóda PACS má vyššiu variabilitu výsledkov a aj značne vyššiu výpočtovú náročnosť. Preto pre ďalšie použitie považujeme ako najlepšieho kandidáta metódu lasso a ako aj iné metódy využívajúce  $l_1$  normy.

## 4.4 Modelovanie popularity

Z prehľadu literatúry sme zistili, že štúdie sa nevenujú čisto vplyvu okolia stanice, ale zahŕňajú aj prediktory reprezentujúce správanie, ako napríklad časy príchodov, ktoré nemusia byť vopred známe pri budovaní staníc. Preto sa zameriavame najmä na vplyv okolia a preskúmavame aj potenciálne ukazovatele výkonnosti nabíjacej infraštruktúry.

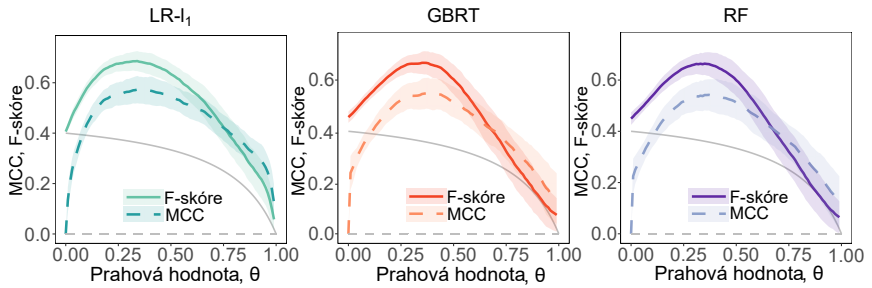
Identifikovali sme niekoľko širších aspektov spojených s nabíjacou infraštruktúrou, ako stimulácia väčšieho užívania EV, smart charging a stabilita distribučných systémov požadujúce súhru medzi elektrickou sieťou a nabíjaním EV a umiestnenie nabíjacích miest na výnosných miestach.

S ohľadom na rozličné uhly pohľadu expertov, samospráv, operátorov elektrických systémov a operátorov nabíjacej infraštruktúry a zvažujúc dostupnosť dát sme na meranie výkonnosti nabíjacích miest vybrali nasledovné ukazovatele, inšpirované [21], uvedené v tabuľke 4.

Indikátory výkonnosti nabíjacích staníc	$R^2$
Spotrebovaná energia [kWh]	0.44
Počet nabíjacích transakcií	0.50
<b>Popularita (počet unikátnych RFID kariet)</b>	<b>0.60</b>
Doba nabíjania [hodiny]	0.48
Pomer nabíjania (doba nabíjania podelená dobou pripojenia)	0.40
Relatívna obsadenosť nabíjacieho miesta (doba nabíjania podelená celkovým časom dostupnosti)	0.42
Relatívne využitie kapacity nabíjacieho miesta (spotrebovaná energia podelená maximálnou menovitou energiou)	0.38

**Tab. 4:**  $R^2$  indikátorov výkonnosti vysvetlených metódou OLS pomocou prediktorov charakterizujúcich okolie a ľudské aktivity v blízkosti nabíjacích miest.

Najvyššiu hodnotu  $R^2$ , a teda najvyšší potenciál pre analýzu dát, má popularita nabíjacích miest (vyjadrená unikátnym počtom RFID kariet, ktorými sa iniciuje nabíjanie), preto sústredíme ďalšie analýzy na tento indikátor. Pri plánovaní umiestnenia novej infraštruktúry je potrebné často vyberať z konečnej množiny kandidátov umiestnenia a v takej situácii netreba odhadnúť presný počet používateľov EV, ale stačí predpovedať, či bude dané nabíjacie miesto populárne. Z tohto dôvodu modelujeme popularitu ako binárnu veličinu, čo redukuje problém, ktorému sa venujeme v tejto podkapitole, na binárnu klasifikáciu. Keďže zvolené metódy vracajú pravdepodobnostnú predpoveď, využívame prahový parameter  $\theta \in \langle 0; 1 \rangle$  pre získanie výsledných binárnych hodnôt. Ak je  $\hat{y} \geq \theta$ , predikcia bude 1, inak 0.



**Obr. 3:** Priemerná hodnota F-skóre a MCC vyhodnotená na testovacích dátach ako funkcia prahovej hodnoty  $\theta$ . Prislúchajúca metóda je uvedená nad grafom. Každá miera je zobrazená iným štýlom čiar. Hrubé čiary odpovedajú priemerným hodnotám pre skupinu 100 rôznych tréningových a testovacích rozdelení dát a plochy reprezentujú jednu smerodajnú odchýlku odhadnutú z dát. Tenké čiary ukazujú hodnoty mier pre nulový model predpovedajúci populárne nabíjacie miesta náhodne s pravdepodobnosťou 0.25.

Pre reprezentáciu okolia sme na základe miery vysvetlenia zvolili polomer kruhovej zóny o veľkosti 350 metrov. Popularitu sme kódovali ako binárnu premennú tak, že sme 25 % najpopulárnejších staníc (najvyšší počet RFID kariet) kódovali ako 1 a zvyšok ako 0.

Na predikcie sme použili metódy GBRT, RF a LR- $l - 1$ , tréňované krížovou validáciou so stratifikáciou [18, s. 70] podľa  $y$ . Predikcie modelov sme evaluovali pomocou mier presnosti, precíznosti a senzitivity [18, s. 256], ako aj hodnotami nulového modelu pre tieto miery pre parameter  $\theta$ . Metódy mali v priemere lepšie výsledky ako nulový model, až na GBRT a RF pre vyššie hladiny  $\theta$ .

Takisto sme vybrali dve doplnkové miery, MCC [22] a F-skóre [32], robustnejšie voči nevyváženosti zastúpenia tried vo vektore odpovedí. Obe miery dosahujú jedno maximum, ktoré je tiež globálne maximum. V tomto maxime metriky ukazujú na zaujímavú kombináciu hodnôt, hlavne pre aplikáciu riešení do praxe.

Obrázok 3 zobrazuje F-skóre a MCC pre všetky tri metódy. Kvalitu predikcie môžeme posúdiť porovnaním výsledkov s nulovým modelom, ktorý predikuje populárne nabíjacie miesta náhodne s pravdepodobnosťou 0.25.

Prahové hodnoty prislúchajúce maximu označujeme ako  $\theta_{MCC_{max}}$  a  $\theta_{F-skóre_{max}}$ . Hodnoty presnosti, precíznosti a senzitivity pre  $\theta_{MCC_{max}}$  a  $\theta_{F-skóre_{max}}$  sú uvedené v tabuľke 5.

Zaujímavé osoby, ako napr. prevádzkovatelia nabíjacej infraštruktúry

	LR- $l_1$	GBRT	RF		LR- $l_1$	GBRT	RF
$\theta_{MCC_{max}}$	0.34	0.35	0.37	$\theta_{F\text{-skóre}_{max}}$	0.34	0.35	0.35
Presnosť	0.83	0.82	0.82	Presnosť	0.83	0.81	0.81
Precíznosť	0.65	0.64	0.63	Precíznosť	0.63	0.61	0.62
Senzitivita	0.73	0.69	0.70	Senzitivita	0.75	0.74	0.73
MCC	0.57	0.55	0.54	F-skóre	0.69	0.67	0.66

**Tab. 5:** Priemerné hodnoty presnosti, precíznosti a senzitivity pre všetky tri metódy odpovedajúce maximálnej hodnote MCC miery a maximálnej hodnote F-skóre.

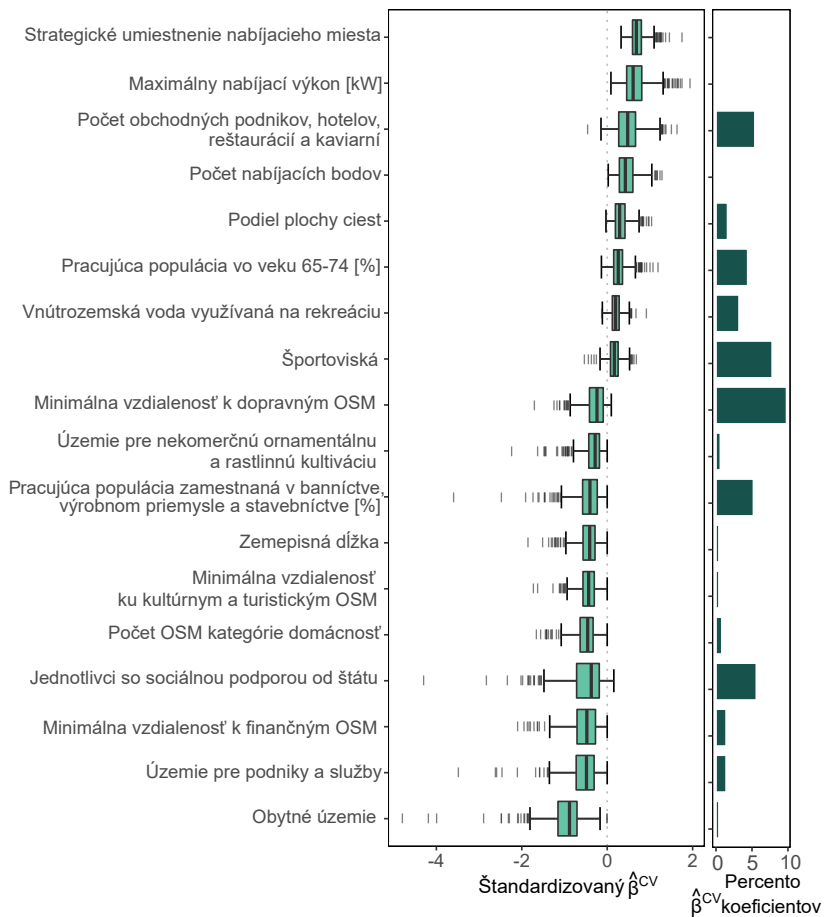
a samosprávy, si vyberú prahovú hodnotu  $\theta$  na základe ich očakávaní a postoja k riziku. Čím nižšia je použitá hodnota  $\theta$ , tým pravdepodobnejšia je úspešná identifikácia populárnych lokácií, pričom sa sa ale zvyšuje riziko umiestnenia nabíjacieho miesta do nepopulárnych oblastí. Naopak čím vyššia je použitá hodnota  $\theta$ , tým vyššia je záruka identifikácie skutočne populárnych nabíjacích miest s nevýhodou možnosti prehliadnutia potenciálne populárnych lokácií. Preto je potrebné zvoliť vhodný kompromis medzi veľkosťou parametra  $\theta$  na základe požadovaného výsledku. Na základe pozorovaných hodnôt mier chybovosti odporúčame prahovú hodnotu  $\theta$  nastaviť v rozsahu od 0.3 po 0.45, kde sú MCC aj F-skóre relatívne vysoké.

Takisto sme aj štatisticky testovali rovnosť rozptylov a priemerov medzi mierou na porovnávanie klasifikácie AUC [18, s. 264] pre jednotlivé metódy, kde LR- $l_1$  malo signifikantne vyššie priemerné hodnoty AUC ako metódy GBRT a RF na hladine významnosti  $\alpha = 0.05$ .

Stromové metódy oproti LR- $l_1$  vrátili príliš komplikované modely zahŕňajúce veľa prediktorov, preto interpretujeme iba výsledky LR- $l_1$  metódy.

Prediktory vybrané pomocou LR- $l - 1$  modelov a štatistickou inferenciou pomocou metódy bootstrap môžu byť kategorizované do troch skupín:

- **Funkcia geografického územia** (v blízkosti nabíjacieho miesta): je signifikantný počet podnikov zameraných na obchod, počet hotelov, reštaurácií a kaviarní, vodné plochy určené na rekreáciu, športové plochy a cesty, všetky s pozitívnym vplyvom. Minimálna vzdialenosť ku kultúrnym a turistickým a finančným OSM objektom mala negatívny koeficient, čo znamená, že čím je taký prediktor vzdialenejší od nabíjacieho miesta, tým klesá popularita nabíjacieho miesta. Ak je takýto objekt v blízkosti nabíjacieho miesta tak môže zvyšovať jeho popularitu.



**Obr. 4:** V ľavej časti grafu je zobrazený Tukey box-plot štandardizovaných koeficientov  $\hat{\beta}_j$  získaný na základe 500  $LR - l_1$  modelov postupom ilustrovaným na obrázku 1. Koeficienty sú usporiadané na základe mediánov ich hodnôt získaných jednotlivými modelmi. Zobrazujeme iba prediktory, ktoré boli vybrané najmenej v 90 % modelov. Stĺpcový graf vpravo zobrazuje percento modelov, v ktorých vyšiel štandardizovaný koeficient  $\hat{\beta}_j$  nulový, t.j.  $j$ -ty prediktor nebol vybratý.

Naopak, obytné územia, územia s nekomerčnou okrasnou a záhradnou kultiváciou a prítomnosť OSM prediktorov súvisiacich s domácnosťou indikujú znižovanie popularity nabíjajúcich miest. Tieto zistenia súvisia s intuíciou, že v obytných územiach sú nabíjacie miesta navštevované homogénnejšou skupinou ľudí ako v rušných urbánnych územiach, pretože v blízkosti obydľí nabíjajú zvyčajne rovnakí používatelia EV, narozdiel od napríklad rušnejších centier miest. Takisto najpravdepodobnejšie vysvetlenie negatívneho vplyvu firemného a industriálneho územia je pracovné nabíjanie [31], t.j. nabíjanie flotily firemných áut alebo zriedkavé využívanie nabíjajúcich miest (malou skupinou) zamestnancov prichádzajúcich do práce.

- **Charakteristiky populácie** (žijúcej v okolí nabíjacieho miesta): Populačná skupina pozitívne spätá s popularitou nabíjajúcich miest sú pracujúci starší ľudia vo veku od 65 do 74 rokov, ktorí sú považovaní za častých vlastníkov EV. Naopak, popularita je negatívne spätá s obyvateľstvom v pracujúcom baníctve, výrobnom priemysle a stavebnom sektore, ako aj s počtom osôb závislých na sociálnej podpore. Tieto výsledky naznačujú, že ekonomická prosperita v blízkosti nabíjajúcich staníc ovplyvňuje návštevnosť nabíjajúcich bodov.
- **Charakteristiky nabíjajúcich miest:** populárne nabíjacie miesta sú pravdepodobne rozmiestnené na základe strategického umiestnenia, majú vyšší maximálny výkon a viac nabíjajúcich bodov. Negatívny vplyv zemepisnej dĺžky môže byť vysvetlený geografiou Holandska, kde je v západnej časti krajiny vyššia urbanizácia keďže sa tam nachádza väčšina veľkých miest.

V tejto podkapitole sme na základe aspektov vyjadrujúcich výkonnosť nabíjacej infraštruktúry definovali sedem indikátorov výkonnosti nabíjajúcich staníc. Ako cieľ sme si stanovili úlohu pokúsiť sa predpovedať popularitu nabíjajúcich staníc, ktorú sme reprezentovali binárnou premennou. Použili sme tri metódy na predikovanie, pričom najlepšie výsledky, no nie výrazne odlišné od ostatných metód, dosiahla metóda  $LR - l_1$ .

## 4.5 Modelovanie energie

Táto podkapitola je primárne venovaná analýze rozdelenia spotreby elektrickej energie medzi jednotlivé nabíjacie miesta pomocou ich okolia, za účelom získania prediktorov vysvetľujúcich toto rozdelenie.

Z dôvodu využitia regresnej metódy lasso a odporúčaní v [s. 92][16] sme si dáta dodatočne predspracovali. Najskôr sme pre zmiernenie multikolinearity iteratívne odstránili prediktory s VIF hodnotou nad 10, transformovali hodnotu  $y$  pomocou  $\log$  transformácie a odstránili hodnoty s nežiadúcou veľkosťou vplyvu idnikovanou Cookovou vzdialenosťou [3, s. 111].

Nakoniec sme po predspracovaní získali 1259 pozorovaní a 119 prediktorov z GIS dát a 5 prediktorov EVnetNL dát. Pri aplikácii OLS metódy na dáta po tomto predspracovaní dostaneme  $R^2 = 0.435$ . Z dôvodu zvýraznenia vplyvu okolia nabíjajúcich miest sme EVnetNL prediktory vynechali a matica  $\mathbf{X}$  bola rozmeru  $n = 1259$ ,  $p = 119$ .

#### 4.5.1 Modelovanie rozdelenia spotreby energie

Pre získanie lepšieho prehľadu o rozdelení spotreby medzi nabíjacie miesta a pre potrebu ďalších štúdií, odhadujeme pravdepodobnostné rozdelenie spotrebovanej energie na nabíjajúcich miestach. Túto energiu modelujeme ako náhodnú premennú  $Y$  s hustotou spotreby  $f_Y(y)$ . Pomocou kombinácie jednoduchých transformácií ( $y$ ,  $y^2$ ,  $y^3$ ,  $\sqrt{y}$ ,  $\sqrt[3]{y}$  a  $\log(y)$ ) a základných rozdelení (Weibullove, beta a gama) sme sa snažili odhadnúť rozdelenie ne-transformovaného vektora  $y$ . Na základe výsledkov Kolmogorov-Smirnov testu, P-P a Q-Q grafov [41] sme zistili, že najlepší odhad poskytuje beta rozdelenie v kombinácii s  $\sqrt[3]{y}$  transformáciou. Výsledný tvar funkcie hustoty rozdelenia  $Y$  je:

$$f(y, \alpha, \beta) = \frac{\left(\frac{\sqrt[3]{y} - y_{min}}{y_{max} - y_{min}}\right)^{\alpha-1} \left(1 - \frac{\sqrt[3]{y} - y_{min}}{y_{max} - y_{min}}\right)^{\beta-1}}{B(\alpha, \beta) 2(y_{max} - y_{min}) y^{\frac{2}{3}}}. \quad (3)$$

Symbol  $B(\alpha, \beta)$  označuje Beta funkciu a odhady parametrov nadobúdajú nasledovné hodnoty:  $\alpha = 2.576$ ,  $\beta = 4.528$ ,  $y_{min} = 91.55$  kWh a  $y_{max} = 16\,649.40$  kWh.

#### 4.5.2 Vysvetľovanie spotreby energie ostatnými indikátormi nabíjajúcich miest

Zaujímavé náhľady na charakteristiky nabíjacej infraštruktúry sme získali analyzovaním vzťahu medzi spotrebou energie a ostatnými charakteristikami nabíjajúcich miest tvoriacich spotrebu energie. Spotreba energie na nabíjacom mieste  $i$  môže byť dekomponovaná na násobok troch iných indikátorov, t.j.

$$y_i = n_i t_i p_i, \quad (4)$$



kde  $n_i$  je počet nabíjácích transakcií, vykonaných na nabíjacom mieste  $i$ ,  $t_i$  je priemerný čas nabíjania na transakciu nabíjacieho miesta  $i$  a  $p_i$  je priemerný nabíjací výkon nabíjacieho miesta  $i$ . Pre všetky nabíjacie miesta tieto kvantily organizujeme ako vektory  $n$ ,  $t$  a  $p$ . Aby sme odhadli, akým spôsobom tieto tri faktory prispievajú k rozdeleniu spotrebovanej energie medzi nabíjacie miesta, skúmame šesť modelov založených na rovnici (4), ktoré vznikli tak, že jeden alebo násobok dvoch indikátorov bol reprezentovaný regresným koeficientom  $k$  odhadnutým OLS bez odstupov, t.j. boli konštantné.

Hlavný faktor spojený s rozdelením spotrebovanej energie medzi nabíjacie miesta je počet transakcií. Fluktuácie v nabíjácích vzoroch (t.j. priemerný nabíjací čas a nabíjací výkon) zohrávajú omnoho menšiu úlohu.

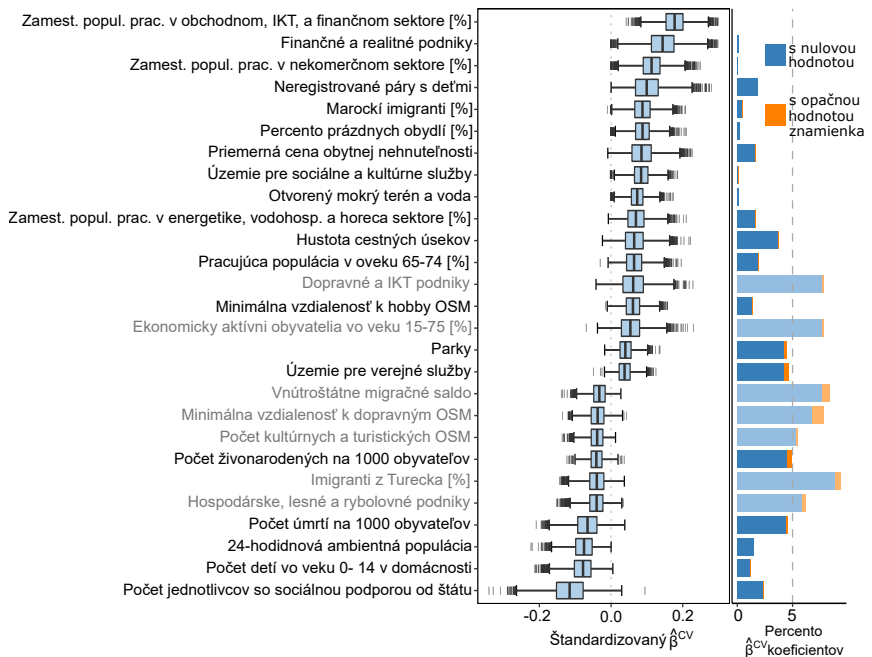
### 4.5.3 Vysvetlenie spotreby energie na nabíjácích miestach pomocou GIS dát

V tejto časti aplikujeme metodológiu spracovania dát na maticu prediktorov  $\mathbf{X}$  odvodenú z GIS dát. Na vektor odpovedí  $y$  (logaritmicky transformovaný) a maticu prediktorov  $\mathbf{X}$  sme aplikovali metódu lasso, trévanú krížovou validáciou, s využitím štatistickej inferencie po výbere premenných pomocou metódy bootstrap, kde používame 10 000 vzoriek.

Zvolené regresné koeficienty môžu byť asociované s tromi rôznymi priestorovými škálami. Niektoré popisujú blízke okolie nabíjácích miest, napr. počet finančných a realitných podnikov, ostatné prislúchajú k obecným samosprávam, napr. prediktory odvodené z datasetu Populačné jadrá, ako napr. percento obyvateľov zamestnaných v nekomerčnom sektore. Posledná skupina regresných koeficientov má potenciál na charakterizovanie lokácie nabíjácích miest na úrovni krajiny, napr. zemepisná šírka a dĺžka.

Regresné koeficienty zobrazené v obrázku 5 sú zoradené zostupne od najvyššej po najnižšiu hodnotu mediánu vzorky. Pre prehľadnosť sme usporiadali signifikantné prediktory s pozitívnym znamienkom mediánu do štyroch skupín:

- **Fyzické prostredie(+):** Územie pre sociálne a kultúrne služby; Otvorený mokrý terén a voda; Hustota ciest; Parky; Územie kde sú poskytované verejné služby.
- **Populácia(+):** Populácia zamestnaná vo veľkoobchode a maloobchode, doprave, skladníctve, IKT službách a finančných službách [%]; Populácia zamestnaná v nekomerčnom sektore [%]; Slobodné páry s deťmi; Imigranti z Maroka [%]; Populácia zamestnaná v energetike,



**Obr. 5:** Empirické rozdelenia štandardizovaných regresných koeficientov získaných lasso metódou kombinovanou s 10 zložkovou krížovou validáciou aplikovanou na 10 000 vzorkách bootstrapovaných dát, získané postupom ilustrovaným na obrázku 1. Zobrazujeme iba prediktory, kde hodnota regresného koeficientu bola nastavená na nulu maximálne v 10 % vzoriek. Koeficienty sú zoradené zostupe od najvyššej hodnoty mediánu po najnižšiu. Ľavý panel zobrazuje Tukeyho box-plot koeficientov. Na pravo je umiestnený vrstvený stĺpcový graf, kde je zobrazené percento vzoriek, kedy bol regresný koeficient  $\hat{\beta}^{CV}$  nastavený na nulu a počet vzoriek dosiahol opačné znamienko ako medián. Za signifikantné považujeme tie prediktory (indikované tmavo modrou farbou), kde je počet vzoriek s nulovým koeficientom menší ako 5% a počet vzoriek s opačným znamienkom je nízky. Svetlo šedá čiarkovaná čiara indikuje 5 % prahovú hodnotu. \*Zamest. popul. prac. - zamestnaná populácia pracujúca.

vodnom a odpadovom hospodárstve a horeca (hotely/reštaurácie/kaviarne) sektore [%]; Pracujúca populácia vo veku 65-74 rokov [%].

- **Služby a podniky(+):** Finančné a realitné podniky; Minimálna vzdialenosť k hobby OSM.
- **Budovy(+):** Percento prázdnych obydlí [%]; Priemerná cena obytnej nehnuteľnosti v tisícoch eur.

Podobne sme organizovali aj signifikantné prediktory s negatívnou hodnotou mediánu:

- **Populácia(-):** Počet jednotlivcov poberajúcich sociálnu pomoc od štátu; Počet osôb vo veku 0 - 14 vo viacčlennej domácnosti s deťmi; Ambientná populácia; Počet úmrtí v roku 2015 na tisíc obyvateľov; Počet živonarodených detí v roku 2015 na tisíc obyvateľov.

Najvyšší počet signifikantných prediktorov sme našli v skupine populácia, k čomu pravdepodobne prispieva aj najvyšší počet prediktorov v tejto skupine. Najväčšia skupina prediktorov indikuje, že vysoký (nízky) príjem a bohatstvo (chudoba) sú pozitívne (negatívne) spojené s množstvom spotrebovanej energie na nabíjajúcich miestach. Najpravdepodobnejšou príčinou sú vysoké ceny EV, čo ich robí dostupné pre lepšie ekonomicky zabezpečených obyvateľov a podniky. Pravdepodobne z toho istého dôvodu sú niektoré signifikantné prediktory spojené s deťmi a mládežou a so staršou populáciou na dôchodku, t.j. so sociálnymi skupinami, ktoré zvyčajne nemajú vysoké príjmy a často sú odkázané na starostlivosť iných osôb. Vysoké percento prázdnych obydlí a vysoká cena obytných nehnuteľností sú asociované s vysokou spotrebou energie na nabíjajúcich miestach. Toto môže reprezentovať novo postavené a nie úplne obývané oblasti s vyšším štandardom života, vyjadreným cez vyššie hodnoty nehnuteľností. Navyše hustota ciest je tiež medzi prediktormi, ktoré sú pozitívne prepojené so spotrebou energie na nabíjajúcich miestach, čo naznačuje, že dobrý prístup k nabíjajúcim miestam môže prispievať k vyššej spotrebe energie.

Po zahrnutí piatich EVnetNL prediktorov (počet nabíjajúcich bodov, maximálny výkon, zemepisná šírka, zemepisná dĺžka a umiestňovacia stratégia) do matice prediktorov, sme analýzu opakovali. Vo všeobecnosti sú signifikantné prediktory podobné ako v obrázku 5, aj keď počet signifikantných prediktorov je nižší, čo pripisujeme nahradeniu niektorých prediktorov EVnetNL prediktormi. Napríklad, mokrý terén a voda, prediktor signifikantný v modeli bez EVnetNL prediktorov, môže byť vyjadrený zemepisnou dĺžkou. Negatívny vplyv zemepisnej dĺžky môže byť vysvetlený

geografickou polohou Holandska, kde je západná časť krajiny viac urbanizovaná a nájdeme tu väčšinu holandských miest. Zároveň je v tejto oblasti veľa povrchovej vody, keďže je zväčša situovaná pod úrovňou mora. Signifikantnosť niektorých prediktorov zo skupín Fyzické prostredie(+) a Populácia(-) bola redukovaná. Všetky EVnetNL prediktory sú významné a maximálny nabíjací výkon, počet nabíjacích bodov a zemepisná dĺžka majú silný vplyv indikujúci potenciálny vplyv parametrov nabíjacích miest na spotrebu energie.

#### 4.5.4 Vplyv stratégie umiestňovania na spotrebu energie

Väčšina nabíjacích miest bola umiestnená použitím jedným z dvoch (strategický alebo dopytovo orientovaný) spôsobov umiestňovania [12]. Strategicky umiestnené nabíjacie miesta sú umiestnené v blízkosti verejných lokalít, kde je intuitívne očakávané nabíjanie EV. Nabíjacie miesta umiestnené dopytovo orientovaným spôsobom sú budované na základe žiadosti od používateľov EV, zvyčajne v blízkosti ich obydli. V tejto sekcii skúmame, či stratégia umiestňovania spôsobuje rozdiel vo faktoroch asociovaných so spotrebou energie. Lasso metódu sme aplikovali na každú skupinu separátne.

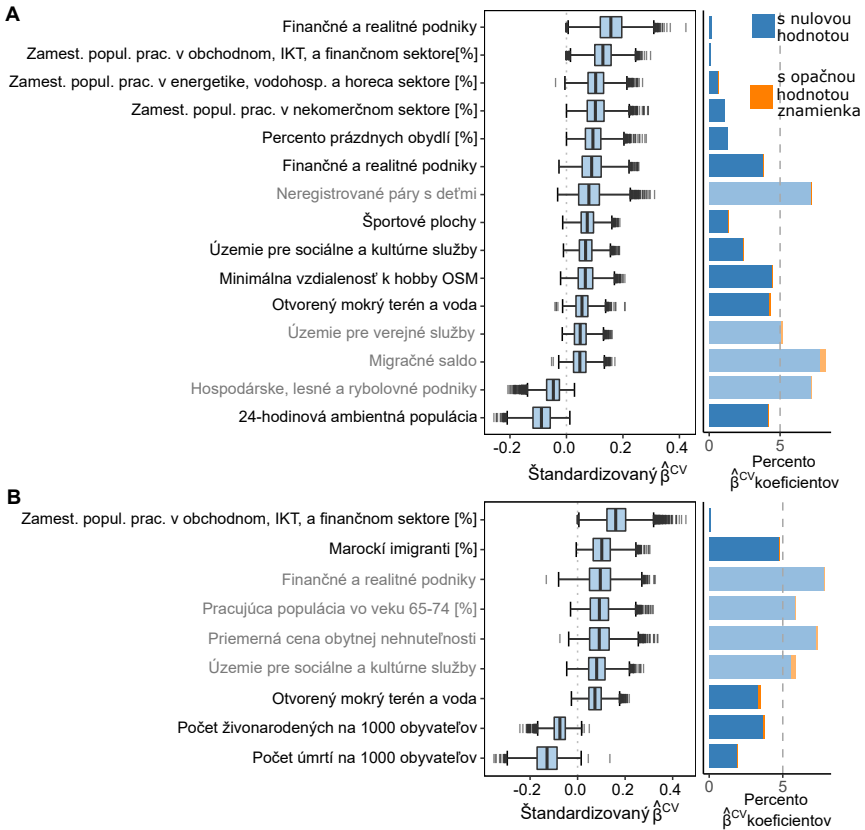
Vybraté prediktory v obrázku 6 sa do veľkej miery zhodujú s prediktormi vybratými pre kompletný dataset (viď obrázok 5).

Elektrická energia spotrebovaná na strategicky umiestnených nabíjacích miestach (obrázok 6A) je pozitívne prepojená s pracovným sektorom obyvateľstva a fyzickým prostredím, t.j. s určitým typom lokalít v blízkosti nabíjacích miest. Pracovné sektory obyvateľov indikujú prevládajúce podniky v obciach a sú pozitívne prepojené so spotrebou energie. Navyše, vybrané prediktory pre strategické umiestňovanie staníc odkazujú na určité podniky a lokácie (športoviská, socio-kultúrne miesta), ktoré môžu byť asociované s príležitostným nabíjaním.

Pre nabíjacie miesta s dopytovo orientovanou stratégiou umiestňovania, (obrázok 6B), negatívny koeficient počtu úmrtí na 1 000 obyvateľov a živonarodených detí na 1 000 obyvateľov indikujú, že územia s vyššou pôrodnosťou a úmrtnosťou sú negatívne spojené so spotrebou energie.

Testovali sme aj ďalšie stratifikácie, podľa ktorých sme delili stanice do neprekrývajúcich sa skupín, napr. na základe počtu nabíjacích bodov, miery obytných zón (reprezentovanej podielom obytného územia v kruhovej zóne nabíjacieho miesta), administratívneho rozdelenia Holandska na 12 provincií a počtu obyvateľov v obciach.

Okrem posledného kritéria, sme obdržali nízky počet vybraných pre-



**Obr. 6:** Zobrazujeme iba prediktory, kde hodnota regresného koeficienta bola nastavená na nulu maximálne v 10 % vzoriek. Empirické rozdelenia štandardizovaných regresných koeficientov získaných lasso metódou kombinovanou s 10 zložkovou krížovou validáciou aplikovanou na 10 000 vzoriek bootstrapovaných dát, získané postupom ilustrovaným na obrázku 1. **A** Strategicky umiestnené nabíjacie miesta. **B** Dopytovo orientované umiestnené nabíjacie miesta. Zobrazujeme iba prediktory, kde hodnota regresného koeficienta bola nastavená na nulu maximálne v 10 % vzoriek. Koeficienty sú zoradené zostupe od najvyššej hodnoty mediánu po najnižšiu. Ľavý panel zobrazuje Tukeyho box-plot koeficientov. Napravo, vrstvený stĺpcový graf ukazuje percento vzoriek, kedy bol regresný koeficient  $\hat{\beta}^{CV}$  nastavený na nulu a počet vzoriek, kedy dosiahol opačné znamienko ako medián. Za signifikantné považujeme tie prediktory (indikované tmavo modrou farbou), kde je počet vzoriek s nulovým koeficientom menší ako 5 % a počet vzoriek s opačným znamienkom je nízky. Svetlo šedá čiarkovaná čiara indikuje 5 % prahovú hodnotu. \*Zamest. popul. prac. - zamestnaná populácia pracujúca.

diktorov. Delením nabíjacích miest na dve skupiny na základe populácie obce, v ktorej sú umiestnené, zvažujúc prahovú hodnotu 50 000 obyvateľov, sme obdržali dve, približne rovnako veľké skupiny. Zaujímavým zistením je, že nabíjacie miesta umiestnené v obciach s viac ako 50 000 obyvateľmi spotrebovali v priemere o 48 % viac elektrickej energie ako nabíjacie miesta druhej skupiny. Vyšší počet signifikantných prediktorov, spojených s populáciou v obciach, finančnými a realitnými podnikmi a fyzickým prostredím, nachádzame vo výsledkoch pre nabíjacie miesta umiestnené v obciach s nižším počtom obyvateľov.

Táto podkapitola sa venovala analýze rozdelenia spotreby elektrickej energie medzi nabíjacie miesta. Spotreba energie má pre nabíjacie miesta po transformácii beta rozdelenie a je najviac ovplyvňovaná počtom transakcií spomedzi základných identifikátorov tvoriacich spotrebu nabíjacích miest. Pri analýze vplyvu okolia sa prejavila najviac ekonomická prosperita obyvateľstva a firiem, či už negatívne alebo pozitívne. Takisto sme ukázali aj faktory vplyvajúce na nabíjacie miesta podľa umiestňovacích stratégií. Tieto faktory môžu slúžiť najmä ako vstup pre optimalizácie umiestnenia nabíjacej infraštruktúry. Ukázaná metodológia je vhodná najmä na identifikáciu faktorov vplývajúcich nielen na spotrebu elektrickej energie, ale aj na iné indikátory, pomocou ich okolia. Takisto ponúka vyhodnotenie signifikantnosti výsledkov po výbere premenných, čo je v literatúre často opomínané.

## Záverčné zhrnutie výsledkov

V tejto práci sa zaoberáme skúmaním možností, ako využiť dátovú analýzu na podporu rozhodovania v oblasti elektromobility.

Pre prvý cieľ sme poskytli riešenie segmentáciou nabíjacích staníc na základe navrhnutých identifikátorov, využitím troch zhlukovacích metód, kde sme porovnávali aj dva prístupy zhlukovania kombinovaného s agregáciou dát. Podarilo sa nám identifikovať 4 zhluky staníc, ktoré sme interpretovali. Lepší z prístupov k zhlukovaniu bol ten, čo najskôr agregoval dáta a následne ich zhlukoval. Zhlukovanie nabíjacích staníc má potenciál prispieť k lepšiemu poznaniu vzorov využívania nabíjacích staníc a môže napomôcť vylepšiť výsledky modelovania indikátorov nabíjacej infraštruktúry pre pokročilejšie učiace sa algoritmy.

Druhý cieľ sme naplnili zostavením, otestovaním a vyhodnotením metód umožňujúcich predpovedať časopriestorovo agregovanú spotrebu nabíjacích staníc. Na základe prechádzajúcich výsledkov a odporúčaní v li-

teratúre sme spotrebu priestorovo agregovali. Uvažovali sme dáta s dennou frekvenciou. Predbežná analýza odhalila v dátach sezónne vzory a aj z toho dôvodu sme využili SARIMAX, GBRT a RF metódy, ktoré sme vylepšili externými prediktormi. Modely sme trénovali tromi procedúrami odlišujúcimi sa v dĺžke trénovacích množín ako aj spôsobe predpovedí. Najlepšie predpovede dosiahla metóda SARIMAX, natrénovaná procedúrou postupne rozširujúcou trénovaciu množinu o pribúdajúce dáta, s presnosťou MAPE okolo 12 %, pričom všetky modely výrazne prekonalí náhodný sezónny model.

Spôsob, akým sme naplnili tretí cieľ, sme opísali v troch podkapitolách. Prvá z týchto podkapitol sa venuje extrakcii prediktorov z GIS dát, reprezentujúcich okolie nabíjajúcich miest a analyzuje schopnosti regresných metód, vyberať premenné v prostredí multikolinearity. Variovali sme stupeň multikolinearity a skúmali, ako sa s ňou jednotlivé metódy vedú vysporiadať. Najlepšie výsledky dosiahli metódy PACS a lasso, pričom metóda lasso mala nižší výpočtový čas a aj nižšiu variabilitu výsledkov. Aj z tohto dôvodu sme metódu lasso a  $l-1$  regularizáciu odporučili pre ďalšie použitie v rámci dizertačnej práce.

Predikovanie popularity nabíjacej infraštruktúry je jedným z hlavných bodov tretieho cieľa. Popularitu sme predikovali na základe charakteristík nabíjajúcich miest a okolia nabíjajúcich miest reprezentovaného GIS dátami. Na základe literatúry a v spolupráci s expertmi v oblasti elektromobility sme identifikovali tri hlavné aspekty pri nasadzovaní a plánovaní nabíjacej infraštruktúry, pomocou ktorých sme vybrali sedem indikátorov výkonnosti nabíjacej infraštruktúry. Medzi nimi mala popularita, kvantifikovaná ako počet unikátnych používateľov EV, čo sa nabíjali na nabíjacom mieste, podstatný význam a najlepšiu vysvetliteľnosť dátami. Popularitu sme pre jednoduchosť a aplikovateľnosť kódovali ako binárnu premennú a využili sme prahovú hodnotu  $\theta$  pre klasifikáciu pravdepodobnostnej hodnoty, s využitím ako miera prístupu k riziku. Okrem metódy LR- $l_1$  sme využili aj stromové metódy RF a GBRT. Všetky použité modely dokázali popularitu predikovať lepšie ako náhodný model, pričom najvyššiu presnosť (0.829) mala metóda LR- $l_1$ . Okrem predikovania sme analyzovali aj vplyvné prediktory, kde sme uprednostnili metódu LR- $l_1$  pred stromovými metódami, ktoré využívali veľký počet prediktorov. Na analýzu vplyvných prediktorov metódy LR- $l_1$  sme využili štatistickú inferenciu pomocou metódy bootstrap. Vplyvné prediktory sme rozdelili do troch skupín: funkcia geografického územia, charakteristiky populácie a charakteristiky nabíjajúcich miest, kde vplyvom dominovali hlavne charakteristiky nabíjajúcich miest.

V poslednej podkapitole, spadajúcej pod tretí cieľ, opisujeme analýzu

závislosti medzi rozdelením spotrebovanej elektrickej energie medzi nabíjacie miesta a charakteristikami okolia nabíjacieho miesta. Najskôr sme analyzovali rozdelenie spotreby energie, ktorá sa dá modelovať po jednoduchej transformácii beta rozdelením. Z indikátorov, ktoré úzko súvisia so spotrebovanou energiou, spotrebu energie najviac ovplyvňuje počet nabíjajúcich transakcií. Pomocou metódy lasso sme identifikovali prediktory okolia nabíjacej infraštruktúry, potenciálne vplyvajúce na spotrebu energie. Využitím prevzorkovacej metódy bootstrap sme indikovali štatistickú spoľahlivosť výsledkov. Prediktory, ktoré významne ovplyvňujú spotrebovanú energiu poukazujú najmä na ekonomickú prosperitu. Napríklad, obyvatelia a firmy s vysokým (nízkym) príjmom, situované v blízkosti nabíjajúcich miest, majú pozitívny (negatívny) vplyv na spotrebu energie. Podobne ekonomická prosperita je vyjadrená drahším novopostaveným bývaním pozitívne prepojeným na spotrebovanú energiu. Takisto významný vplyv má pracovný sektor obyvateľstva v samosprávach, ako aj počet finančných a realitných podnikov. Najväčší negatívny vplyv majú obyvatelia závislí na sociálnej pomoci. Stratifikácia nabíjajúcich miest, podľa stratégie použitej pri jej umiestňovaní, viedla k rozdeleniu regresných koeficientov na dve skupiny. Typy podnikov, pracovný sektor obyvateľstva a verejné miesta záujmu sú prepojené na vyššiu spotrebu energie na nabíjajúcich miestach umiestnených strategicky. Populačné charakteristiky, napr. počet živých pôrodov a úmrtí na 1 000 obyvateľov, sú prepojené na spotrebu energie na nabíjajúcich miestach umiestnených na základe dopytu.

Primárne využitie výsledkov vidíme v podpore pri zostavovaní optimalizačných, simulačných, ako aj učiacich sa modelov na základe nami vybratých dát, vhodne popisujúcich okolie nabíjajúcich staníc. Takto získané modely dokážu pomôcť efektívnejšie rozmiestniť nabíjajúcu infraštruktúru a prispieť tak k energetickej efektívnosti a ekologickejšej doprave. Zostavená metodológia na stratifikáciu staníc, popísaná v podkapitole 4.5, môže byť použitá na vylepšenie umiestňovacích stratégií nabíjacej infraštruktúry, zameraných na určité skupiny nabíjajúcich staníc, napr. stanice určené na pracovné nabíjanie. Takisto zostavená metodológia v podkapitolách 4.2 a 4.5 môže byť použitá na plánovanie kapacity elektrickej siete a to zvážením energetického dopytu nabíjacej infraštruktúry.

Za hlavné teoretické ako aj praktické prínosy pokladáme výsledky dátových analýz, ktoré priniesli nové poznatky o správaní sa obslužného systému, skladajúceho sa s veľkého množstva nabíjajúcich staníc. Boli to napríklad pravdepodobnostné rozdelenie spotreby energie, charakteristiky časových radov, správanie popísané jednotlivými zhlukmi nabíjajúcich staníc, ako aj samotné interpretácie koeficientov, ktoré odpovedajú prediktorom



potenciálne vplývajúcim na popularitu a spotrebu energie. Takisto medzi dôležité prínosy práce patrí porovnanie metód, ktoré bolo možné použiť pri riešení jednotlivých problémov (napr. predpovedanie časových radov, určenie zhlukov staníc, predikcia popularity atď.). Určili sme, ktorá z uvažovaných metód alebo postupov je pre daný problém najvhodnejšia.

Na základe tohoto súhrnu považujeme ciele práce za splnené. Uvedené výsledky výskumu prinášajú nové možnosti aplikácie a nové poznatky pre vednú oblasť inteligentných informačných systémov a spolu s použitou metodológiou môžu nájsť uplatnenie v oblasti elektromobility.

### **Odporúčania pre ďalšiu prácu**

V budúcnosti by sme sa chceli zamerať na tvorbu segmentov zameranú na časové využívanie nabíjacej infraštruktúry. V porovnaní s [13] by sme chceli dosiahnuť spresnenie predpovedania časov pripojenia k nabíjacím staniciam. Zaujímavé výsledky môžu priniesť aj zopakovanie analýzy identifikátorov vplývajúcich na prevádzku nabíjacej infraštruktúry samostatne pre veľké mestské zóny ako Amsterdam alebo Rotterdam. Takisto ďalej spolupracujeme s autormi [37] na pravdepodobnostných predikciách spotreby energie s hodinovou frekvenciou.

# Publikácie autora v skúmanej oblasti

**ADC** - Vedecké práce v zahraničných karentovaných časopisoch

Straka, Milan; De Falco, Pasquale; Ferruzzi, Gabriella; Proto, Daniela; van der Poel, Gijis; Khormali, Shahab; Buzna Luboš: "Predicting popularity of electric vehicle charging infrastructure in urban context", *IEEE Access* (IF 4.098), s. 11315-11327, 2020, ISSN 2169-3536

**AFC** - Publikované príspevky na zahraničných vedeckých konferenciách

Buzna, Luboš; De Falco, Pasquale; Khormali, Shahab; Proto, Daniela; Straka, Milan: "Electric vehicle load forecasting : a comparison between time series and machine learning approaches", *1st International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED 2019)*, Danvers : Institute of Electrical and Electronics Engineers, s. 1-5, 2019, ISBN 978-1-7281-3088-0

Straka, Milan; Buzna, Luboš: "Use case and introductory analysis of the dataset collected within the large network of public charging stations"; *Reliability and statistics in transportation and communication: selected papers from the 18th international conference*, Cham : Springer Nature, s. 203-213, 2019, ISBN 978-3-030-12449-6

**AFD** - Publikované príspevky na domácich vedeckých konferenciách

Straka, Milan: "Robustness of variable selection regression methods", *Proceedings of the MIST conference 2018*, CreateSpace Independent Publishing Platform, s. 1-6, 2018, ISBN 9781987603392

Straka, Milan; Buzna, Luboš: "Clustering algorithms applied to usage related segments of electric vehicle charging stations" *TRANSCOM 2019: conference proceedings*, Amsterdam : Elsevier Science, s. 1576-1582, 2019, ISSN 2352-1465, *Transportation Research Procedia*, Vol. 40. - ISSN 2352-1465

Straka, Milan; Buzna, Luboš: "Preprocessing of GIS data for electric vehicle charging stations analysis and evaluation of the predictors significance", *TRANSCOM 2019 : conference proceedings*, Amsterdam : Elsevier Science, s. 1583-1590, 2019, ISSN 2352-1465, *Transportation Research Pro-*

## Zoznam použitej literatúry

- [1] Berk, R.; Brown, L.; Buja, A.; et al.: Valid post-selection inference. *The Annals of Statistics*, volume 41, no. 2, 2013: pp. 802–837.
- [2] Bikcora, C.; Refa, N.; Verheijen, L.; et al.: Prediction of availability and charging rate at charging stations for electric vehicles. In *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct 2016, p. 1–6, doi:10.1109/PMAPS.2016.7764216.
- [3] Chatterjee, S.; Hadi, A. S.: *Regression analysis by example*. John Wiley & Sons, 2015.
- [4] Cuesta, H.; Kumar, S.: *Practical Data Analysis*. Packt Publishing Ltd, 2016.
- [5] Dannecker, L.: *Energy time series forecasting: efficient and accurate forecasting of evolving time series from the energy domain*. Springer, 2015.
- [6] Economist, T.: The world's most valuable resource is no longer oil, but data. *The Economist: New York, NY, USA*, 2017.
- [7] ElaadNL: ElaadNL. <https://www.elaad.nl/>, 2018, accessed: 2019-06-21.
- [8] EnergieAtlas, N.: Energy Atlas. <https://www.pdok.nl/downloads?articleid=1951681>, 2015, accessed: 2018-10-16.
- [9] GreenWay: GreenWay. <https://greenway.sk>, 2018, accessed: 2019-06-24.
- [10] Han, J.; Pei, J.; Kamber, M.: *Data mining: concepts and techniques*. Elsevier, 2011.
- [11] Hastie, T.; Tibshirani, R.; Friedman, J.: *The elements of statistical learning: data mining, inference and prediction*. Springer, second edition, 2009.  
URL <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- [12] Helmus, J.; Spoelstra, J.; Refa, N.; et al.: Assessment of public charging infrastructure push and pull rollout strategies: The case of the Netherlands. *Energy policy*, volume 121, 2018: p. 35–47.
- [13] van den Hoed, R.; Maase, S.; Helmus, J.; et al.: *E-mobility getting smart with data*. Amsterdam University of Applied Sciences, 2019.  
URL [https://www.researchgate.net/publication/334625203\\_E-mobility\\_getting\\_smart\\_with\\_data](https://www.researchgate.net/publication/334625203_E-mobility_getting_smart_with_data)
- [14] Hsu, D.: Identifying key variables and interactions in statistical models of building energy consumption using regularization. *Energy*, volume 83, 2015: pp. 144–155.
- [15] Hyndman, R. J.; Athanasopoulos, G.: *Forecasting: principles and practice*. OTexts, 2018.

- [16] James, G.; Witten, D.; Hastie, T.; et al.: *An introduction to statistical learning*, volume 112. Springer, 2013.
- [17] Kotka, V.; Koreň, M.; Ižvoltová, J.: *Geoinformatika pre inžinierov*. Edis, Žilina, 215.
- [18] Kuhn, M.; Johnson, K.: *Applied predictive modeling*, volume 26. Springer, 2013.
- [19] Liu, S. V.; Chen, F.-L.; Xue, J.: Evaluation of Traffic Density Parameters as an Indicator of Vehicle Emission-Related Near-Road Air Pollution: A Case Study with NEXUS Measurement Data on Black Carbon. *Int J Environ Res Public Health*, volume 14, no. 12, 2017: p. 1581.
- [20] Louie, H. M.: Time-Series Modeling of Aggregated Electric Vehicle Charging Station Load. *Electric Power Components and Systems*, volume 45, no. 14, 2017: pp. 1498–1511, doi:10.1080/15325008.2017.1336583, <https://doi.org/10.1080/15325008.2017.1336583>. URL <https://doi.org/10.1080/15325008.2017.1336583>
- [21] Lucas, A.; Prettico, G.; Flammini, M. G.; et al.: Indicator-Based Methodology for Assessing EV Charging Infrastructure Using Exploratory Data Analysis. *Energies*, volume 11, no. 7, 2018, ISSN 1996-1073, doi: 10.3390/en11071869. URL <http://www.mdpi.com/1996-1073/11/7/1869>
- [22] Matthews, B. W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, volume 405, no. 2, 1975: p. 442–451.
- [23] Ministry of the Interior and Kingdom Relations: Liveability Meter. <https://data.overheid.nl/data/dataset/leefbaarometer-2-0---meting-2016>, 2015, accessed: 2018-10-15.
- [24] Mitchell, T. M.; et al.: Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, volume 45, no. 37, 1997: pp. 870–877.
- [25] Netherlands Enterprise Agency: Electric vehicle charging - Definitions and explanation. <https://bit.ly/2LehwPk>, 2019, accessed: 2019-06-19.
- [26] NRSR: Zákon o energetike a o zmene a doplnení niektorých zákonov. <https://www.zakonypreludi.sk/zz/2012-251>, 2019, accessed: 2019-12-15.
- [27] Oak Ridge National Laboratory: LandScan Datasets. <https://landscan.ornl.gov/landscan-datasets>, 2015, accessed: 2015-05-20.
- [28] OpenChargeMap: <https://openchargemap.org>, 2015, accessed: 2019-01-10.
- [29] OpenStreetMap: <https://www.openstreetmap.org>, 2015, accessed: 2019-02-13.
- [30] Oplaadpalen: <https://www.oplaadpalen.nl/>, 2015, accessed: 2019-02-20.

- [31] Sadeghianpourhamami, N.; Refa, N.; Strobbe, M.; et al.: Quantitive analysis of electric vehicle flexibility: A data-driven approach. *International Journal of Electrical Power & Energy Systems*, volume 95, 2018: p. 451–462, ISSN 0142-0615, doi:10.1016/j.ijepes.2017.09.007.  
URL <http://www.sciencedirect.com/science/article/pii/S0142061516323687>
- [32] Sasaki, Y.; et al.: The truth of the F-measure. 2007.
- [33] Sharma, D. B.; Bondell, H. D.; Zhang, H. H.: Consistent group identification and variable selection in regression with correlated predictors. *Journal of Computational and Graphical Statistics*, volume 22, no. 2, 2013: pp. 319–340.
- [34] Statistics Netherlands: CBS Land Cover. <https://www.pdok.nl/downloads?articleid=1951731>, 2015, accessed: 2016-09-14.
- [35] Statistics Netherlands: Neighbourhoods dataset 2015. <https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/wijk-en-buurtkaart-2015>, 2015, accessed: 2018-08-20.
- [36] Statistics Netherlands: Population cores in the Netherlands. <https://www.cbs.nl/nl-nl/achtergrond/2014/13/bevolkingskernen-in-nederland-2011>, 2015, accessed: 2017-09-14.
- [37] Straka, M.; De Falco, P.; Ferruzzi, G.; et al.: Predicting Popularity of Electric Vehicle Charging Infrastructure in Urban Context. *IEEE Access*, volume 8, 2020: pp. 11315–11327.
- [38] Thornton, L. E.; Pearce, J. R.; Kavanagh, A. M.: Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. *International Journal of Behavioral Nutrition and Physical Activity*, volume 8, no. 1, 2011: p. 71.
- [39] Tibshirani, R.; Wainwright, M.; Hastie, T.: *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC, 2015.
- [40] Traffic flows: The database was provided for reserach purposes by the National Institute for Public Healt and Environment. <http://www.rivm.nl/>, 2015, accessed: 2019-01-07.
- [41] Warren-Hicks, W. J.; Hart, A.: *Application of uncertainty analysis to ecological risks of pesticides*. CRC Press, 2010.
- [42] Xydas, E.; Marmaras, C.; Cipcigan, L. M.; et al.: A data-driven approach for characterising the charging demand of electric vehicles: A UK case study. *Applied energy*, volume 162, 2016: pp. 763–771.