

Analýza dát zo cyklosčítačov v Bratislave

Tím: Jednotné programovacie družstvo a.s.

Lucia Ganajová
Tuan Dávid Nguyen Van
Jakub Novotný
Marek Šugár

Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
2026

Úvod

V posledných rokoch je z hľadiska urbánneho plánovania badať tendencie znižovania akejsi nadvlády áut v uliciach. Práve naopak, najmä v krajských mestách je zvyšovaný dôraz na budovania infraštruktúr pre chodcov a cyklistov. Výnimkou nie je ani Bratislava, v ktorej v posledných rokoch je možné pozorovať nárast práve prepravy po dvoch kolesách.

Magistrát mesta aj z dôvodu ďalšej analýzy inštaloval od roku 2014 na niektorých úsekoch cyklotrás sčítače, ktoré s hodinovou frekvenciou akumulujú absolútne početnosti cyklistov, ktorí v danom časovom intervale prešli ktorýmkoľvek z dvoch smerov. Toto veľké množstvo dát vytvára priestor na hlbšie analýzy, ktoré bližšie dokážu opísať správanie obyvateľov Bratislavy.

V tomto reporte je cieľom poukázať na zaujímavé interpretačné dôsledky, vychádzajúce z dát. Dôraz je kladený interpretáciu dát s rôznou frekvenciou a spojitosti, resp. rozdielnosti medzi jednotlivými cyklotrasami. Okrem iného, hlbší dôraz je kladený spojitost celkovej vyťaženia cyklotrás s počasím.

Report vznikol ako jeden z podkladov riešenia semestrálneho projektu v rámci predmetu **Princípy dátovej vedy** na zimnom semestri 2025-2026 na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

v Bratislave, január 2026
autori

Obsah

1	Spojitosť vyťaženia cyklotrás a počasia	3
1.1	Predspracovanie dát	3
1.2	Lineárna regresia	4
1.3	Penalizované lineárne regresie	5

Kapitola 1

Spojitosť vyťaženia cyklotrás a počasia

Pre účely bližšieho porozumenia vplyvu stavu počasia v Bratislave na celkový počet absolútnych prejazdov po jednotlivých cyklotrasách je možné zvoliť viacero možných prístupov. V projekte bol zvolený prístup tréningu vybraných algoritmov strojového učenia a následnej evaluácie predikovania modelov.

1.1 Predspracovanie dát

Dostupné dátové zdroje poskytujú s hodinovou frekvenciou informácie o počte prejazdov v oboch smeroch na danej cyklotrase. Dátové zdroje akumulujúce údaje o stave počasia boli avšak dostupné v spoľahlivej podobe iba s dennou frekvenciou.

Z tohoto dôvodu bolo vhodné agregovať dáta z jednotlivých dostupných hodín prejazdov do konkrétnych dní. Pre účely priamočiarejšej a úspornejšej interpretácie má zmysel agregovať aj dáta z oboch smerov cyklotrasy dokopy – tým získavame počet denných prejazdov na cyklotrase celkovo. Pre účely interpretácie vplyvu počasia je to z apriórneho hľadiska postačujúce, z hľadiska dostupných dátových zdrojov jedna z mála možností.

Počasia dátový zdroj zrkadlí pomocou viacerých čiastkových premenných, niektoré viacej aj menej interpretačne uchopiteľné. Z relatívne bohatého dátového zdroja je možné vybrať reprezentatívnu podvzorku premenných – využívame prístup interpretačnej spojitosti na základe istého poznania správania.

Do našej využitej množiny premenných zaraďujeme údaje o teplote – **priemerná teplota**, **minimálna** a **maximálna teplota**. Na základe týchto dát je možné očakávať priamu úmeru s absolútnym počtom prejazdov cyklotrasou. Vyššie teploty počas dňa pravdepodobne indukujú vyšší počet prejazdov, vyššiu tendenciu ľudí uprednostniť tento dopravný prostriedok pred napr. *autom*, *MHD* a pod.

Významný vplyv na vyťaženosť cyklotrasy a na s ním prepojený sentiment obyvateľov voči cyklo-doprave má určite **spád zrážok**. Do databázy premenných boli vložené obe dostupné metriky spádu zrážok – spád dažďových a snehových zrážok.

Okrem iného významnými metrikami sú **rýchlosť vetra** a **tlak**. Prvá spomenutá metrika má dozaista vplyv tendenciu ľudí uprednostniť dopravu bicyklom, spoločne s tlakom vzduchu, ktorý má významný vplyv na vyvíjanú biozáťaž a tendenciu ľudí vykonávať fyzickú záťaž na ráme, takpovediac,

nutnosti.

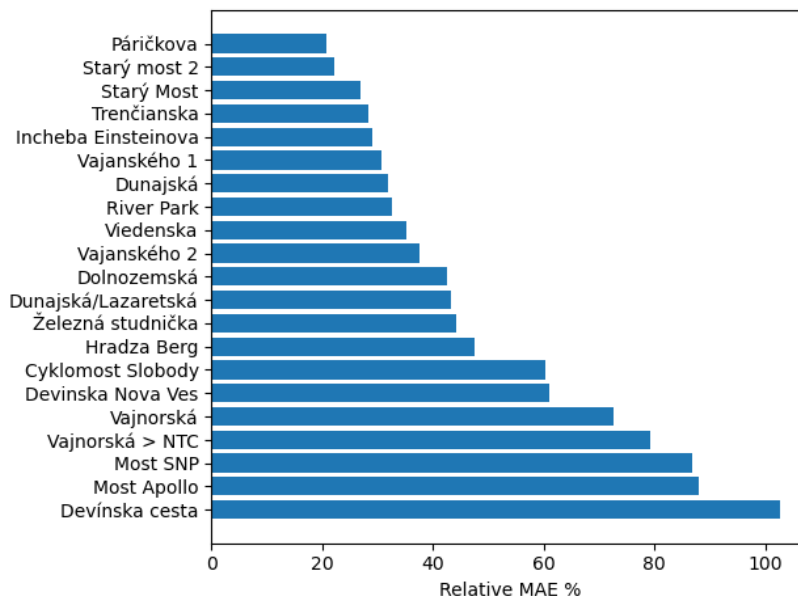
Pri delení dát do zodpovedajúcich tréningových, validačných a testovacích množín dát je nutné zachovať chronologickú následnosť dát, keďže v podstate sa jedná o časové rady a neuváženým prístupom k deleniu je možné disponovať skreslenými, zväčša nadhodnotenými, výsledkami výkonnosti modelov.

Z dôvodu variabilnej dĺžky monitorovaného časového intervalu pre rôzne cyklotrasy nie je možné pristupovať k delbe pomocou fixných časových breakpointov. V projekte bol využitý prístup rozdelenia celkového intervalu v pomere 6 : 3 : 1 pre tréningovú, validačnú a testovaciu množinu. Postupnosť dátových vstupov je zabezpečená zachovaním vzájomnej následnosti jednotlivých podintervalov. Tréningová množina predstavuje prvých 60% dát, validačná nasledujúcich 30% a zvyšok testovacia množina.

1.2 Lineárna regresia

Prvotnou ideou spomedzi klasických prístupov strojového učenia je lineárna regresia. Na tomto mieste je vhodné ozrejmiť aj metriku úspešnosti modelovania, na ktorú budeme výkonnosti modelov reflektovať. V kapitole je využitá **priemerná absolútna odchýlka**, normalizovaná zodpovedajúcim priemerom ako $\frac{MAE(Y, \hat{Y})}{\bar{Y}}$, kde Y označuje skutočnú hodnotu prejazdov, \hat{Y} predikcia modelu a \bar{Y} priemer skutočných hodnôt. Normalizácia je vhodná najmä z dôvodu relativizácie presnosti predikcie medzi rôznymi cyklotrasami. Metrika **priemernej absolútnej percentuálnej odchýlky (MAPE)** sa preukázala byť zväčša numericky nestabilná, pri veľmi nízkych skutočných počtoch prejazdov a i napriek tomu relatívne presnej predikcii metriky indikovala abnormálne vysoké hodnoty.

Výkonnosti modelovania pre jednotlivé cyklotrasy sú vizualizované na *Obrázku 1*.



Obr. 1.1: Presnosť modelovania počtu prejazdov cyklotrasou lineárnou regresiou.

Je možné pozorovať pomerne široké spektrum presností - nižšie aj vyššie odchýlky.

1.3 Penalizované lineárne regresie