

System pro klasifikaci dat kriminality města Most

Dan Friedjung*

Abstrakt

Kriminalita a pocit bezpečí patří mezi klíčová témata, která ovlivňují kvalitu života obyvatel měst. V dnešní době máme díky digitálním technologiím a dostupnosti dat nové možnosti, jak porozumět bezpečnostní situaci ve městech lépe a v širších souvislostech. Tato práce představuje nástroj pro periodickou klasifikaci incidentů z města Most pomocí Python skriptu, který na základě textových popisů automaticky přiřazuje případy do předem definovaných tříd kriminality. Skript zároveň slouží ke sjednocení dat ze dvou různých zdrojů – databází Městské policie Most a Policie České republiky – a převádí je do jednotné struktury pro další analýzu. Závěrečnou částí práce je vizualizace a prostorová analýza, která ukazuje, jak lze s klasifikovanými daty dále pracovat. Výsledná data díky své struktuře a kvalitě otevírají široké možnosti pro další využití.

*xfried08@vutbr.cz, Faculty of Information Technology, Brno University of Technology

1. Úvod

Každý z nás už asi někdy něco hledal na interaktivní mapě – třeba kde je plánovaná odstávka elektřiny, kam zajít na oběd nebo kudy se vyhnout zácpě. Takové mapy nám pomáhají zorientovat se ve světě kolem nás díky datům.

Tato práce se zaměřuje na vytvoření Python skriptu, který periodicky klasifikuje data o kriminalitě ve městě Most na základě textových popisů incidentů. Skript převádí záznamy z databází Městské policie Most a Policie ČR do jednotné struktury, což umožňuje jejich další analýzu a vizualizaci například v interaktivní mapové aplikaci. Klasifikovaná data tak poskytují ucelený přehled o bezpečnostní situaci ve městě a umožňují efektivní monitorování a plánování.

2. Cíle práce

Cílem této práce bylo navrhnout a implementovat Python skript, který by městu Most umožnil periodicky klasifikovat textové záznamy trestných činů. Součástí řešení byl také návrh vhodného datového modelu pro uchování klasifikovaných dat a schopnost propojit tyto záznamy s daty od Policie ČR. To vše tak, aby byl skript skutečně použitelný v praxi a byl tedy jednoduše konfigurovatelný.

Následně bylo žádoucí, také využít klasifikovaná data

k vytvoření jednoduché webové mapy, která by sloužila k jejich vizualizaci a základní analýze. Aplikace měla umožnit interaktivní prohlížení incidentů v prostoru města Most, včetně možnosti filtrování podle kategorií nebo časového období. Cílem bylo ukázat, že klasifikovaná data mohou být snadno využita v praxi a integrovatelná do nástrojů pro podporu rozhodování.

3. Klasifikační modely

Skript podporuje volbu ze sedmi základních klasifikačních modelů, které jsou definovány v konfiguračním souboru. Mezi tyto modely patří: SVM (Support Vector Machine), Naive Bayes (NB), Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting Decision Trees (GBDT), Logistic Regression (LR) a Multi-Layer Perceptron (MLP) [1]. Na vodorovném grafu (2) je zobrazeno porovnání úspěšnosti těchto modelů na finální testovací a trénovací sadě dat. Z výsledků je patrné, že nejlepší výkony vykazují modely SVM, GBDT a MLP, které dosahují přesnosti klasifikace přes 94%. Vzhledem k tomu, že vstupní data jsou často velmi nestrukturovaná a rozmanitá, lze tyto hodnoty považovat za nadstandardní a velmi dobrý výsledek. Zároveň ale platí, že právě velká různorodost dat může vést ke zkreslení finální přesnosti, a proto je vždy důležité jednotlivé modely porovnávat a sledovat vývoj jejich výkonnosti v čase.

Aby bylo možné klasifikační modely efektivně vytrénovat, bylo nejprve nutné manuálně klasifikovat trénovací datovou sadu. Tato sada se postupně rozšiřovala o další ručně zpracované záznamy, čímž se zlepšovala přesnost trénovaných modelů. Z grafu (3) je pak patrné, jak se s růstem objemu trénovací skupiny dat postupně zvyšovala přesnost modelů LR a GBDT, což potvrzuje přínos rozšiřování trénovacích dat pro výkon těchto algoritmů.

4. Webová aplikace

Byla vytvořena interaktivní webová mapa, která vizualizuje klasifikovaná data o přestupcích v městě Most. Při vysoké hustotě incidentů se body automaticky shlukují do clusterů, jež se při přiblížení rozpadnou zpět na jednotlivé záznamy.

Kliknutím na libovolný bod se otevře popup s detaily o daném přestupku – zobrazí se třídy, podtřídy, datum, zdroj a další relevantní informace viz. (4).

Uživatelé mohou filtrovat zobrazené události podle tříd a podtříd kriminality či podle časového období, což umožňuje rychlou selekci vybraných typů incidentů.

Mapa je momentálně nasazena a testována výhradně v interním prostředí města Most, aby bylo zajištěno bezpečné nakládání s daty.

5. Analýza

Nad klasifikovanými daty byla provedena základní analýza pomocí grafů i heatmap. Graf (5) znázorňuje postupný pokles celkového počtu deliktů v Mostě v letech 2013 až 2025.

Heatmapy (6) pak umožňují srovnat intenzitu přestupků spojených s alkoholismem a toxikomanii v centru města mezi lety 2015 a 2023 – ve všech vymezených zónách je patrný výrazný úbytek těchto incidentů. Tento trend lze pravděpodobně přičíst vyhlášce z roku 2017, která zpřísnila pravidla pro konzumaci alkoholu na veřejných prostranstvích.

Kromě uvedených trendů, může podrobná práce s klasifikovanými daty odhalit řadu dalších zajímavých korelací a souvislostí. Právě díky předchozí klasifikaci lze tyto pokročilé analýzy provádět rychle a efektivně, čímž se otevírá prostor pro hlubší porozumění dynamice kriminality v městě Most.

Na přiloženém QR kódu najdete odkaz na mou StoryMapu, která některé základní analýzy rozebírá podrobněji v rámci soutěže Student GIS Talent.

6. Testování

Testování probíhalo na datech poskytnutých městem Most, která sloužila pro ověření funkčnosti a přesnosti klasifikace. Následně byl skript nasazen přímo na straně města do GIS stacku, kde byl provozován v reálném prostředí, tak jak bude využíván v běžném provozu.

Na základě výstupů tohoto testování byly upraveny některé klíčové funkcionality – zejména konfigurace parametrů běhu, způsob zpracování logů a optimalizace výkonu při větších objemech dat.

7. Závěr

V rámci práce se podařilo vytvořit funkční nástroj pro klasifikaci a vizualizaci incidentů z města Most. Hlavním výstupem je Python skript pro periodické zpracování záznamů Městské policie, který umožňuje také jejich spojení s daty Policie ČR. Součástí řešení je datový model pro uchování klasifikovaných dat a jednoduchá webová aplikace pro jejich vizualizaci a základní analýzu. Celý systém je přizpůsoben pro praktické nasazení v prostředí města Most.

Poděkování

Rád bych poděkoval městu Most za poskytnutí dat, panu Ing. Kamilu Novákovi z GIS oddělení za vstřícné konzultace a cenné rady během zpracování práce, a také panu Ing. Jiřímu Hynkovi, Ph.D., za vedení bakalářské práce a odbornou podporu.

Literatura

- [1] Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. Text classification algorithms: A survey. *Information*, 10(4), 2019.