



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INFORMAČNÍCH SYSTÉMŮ

DEPARTMENT OF INFORMATION SYSTEMS

ANALÝZA DAT ZE SLUŽBY PRO SDÍLENÍ KOL VE MĚSTĚ BRNO

ANALYSIS OF DATA FROM THE BIKE SHARING SERVICE IN BRNO

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

VEDOUCÍ PRÁCE

SUPERVISOR

DAVID SKŘEČEK

Ing. JURAJ LAZÚR

BRNO 2025

Zadání bakalářské práce



161039

Ústav: Ústav informačních systémů (UIFS)
Student: **Skřeček David**
Program: Informační technologie
Název: **Analýza dat ze služby pro sdílení kol ve městě Brno**
Kategorie: Webové aplikace
Akademický rok: 2024/25

Zadání:

1. Prostudujte problematiku chytrých měst s důrazem na plánování a provoz v oblasti cyklistické dopravy. Zaměřte se na služby sdílení kol.
2. Prostudujte problematiku zpracování, analýzy a vizualizace dat.
3. Seznamte se s údaji o využití a obsazenosti stanic pro sdílení kol v Brně a s údaji o cyklistické infrastruktuře města. Analyzujte požadavky města na vyhodnocování a vizualizaci těchto dat.
4. Navrhněte systém pro vyhodnocování a vizualizaci dat z bodu 3. Zaměřte se na vyhodnocování území z hlediska kritických míst (přebytek/nadbytek kol), predikujte potenciál poptávky/nabídky v závislosti na různých aspektech (např. počasí, sklonu terénu).
5. Navržený systém implementujte.
6. Výsledné řešení otestujte ve spolupráci s městem Brno.

Literatura:

- Cepero, Teresa, et al. "Visualization in smart city technologies." *Ibero-American Congress of Smart Cities*. Springer, Cham, 2022
- Moustaka, Vaia, Athena Vakali, and Leonidas G. Anthopoulos. "A systematic review for smart city data analytics." *ACM Computing Surveys (cSuR)* 51.5 (2018): 1-41
- Freund, Daniel, Shane G. Henderson, and David B. Shmoys. "Bike sharing." *Sharing Economy: Making Supply Meet Demand* (2019): 435-459
- Li, Yexin, et al. "Traffic prediction in a bike-sharing system." *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*. 2015
- Oliveira, Guilherme N., et al. "Visual analysis of bike-sharing systems." *Computers & Graphics* 60 (2016): 119-129

Při obhajobě semestrální části projektu je požadováno:
Body 1. až 4.

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Lazúr Juraj, Ing.**
Vedoucí ústavu: Kolář Dušan, doc. Dr. Ing.
Datum zadání: 1.11.2024
Termín pro odevzdání: 14.5.2025
Datum schválení: 22.10.2024

Abstrakt

Sdílená městská mobilita je stále rozvíjející se fenomén, který, jako pouze další z mnoha, využívá k provozu svých služeb hojné množství mobilních telefonů. Raná fáze rozvoje tohoto odvětví ponechává prostor vylepšením, například pomocí pokročilé datové analýzy, například použitím prediktivních modelů, jež mezi nástroji pro analýzu sdílené cyklistiky nemají zastoupení. Tyto modely však umožňují důležitou formu analýzy, která by uživatelům umožnila plánovat své cesty na základě pravděpodobné dostupnosti dopravních prostředků. Cílem této práce bylo vytvořit komplexní analytický nástroj sdílené cyklistické dopravy ve městě Brno, který umožní zjištění dostupnosti v budoucnu. Prediktivní model, trénovaný na historických datech provozovatelů služeb sdílené cyklistiky, umožňuje s vysokou spolehlivostí predikovat míru pravděpodobnosti výskytu dopravního prostředku na vybrané stanici v požadovaném datu a čase. Výsledkem je nejen tímto způsobem analyzující nástroj, ale také poznatky o potřebách a limitech bránících spolehlivé predikční analýze. Řešení poskytuje inovativní přístup k analýze dat sdílené cyklistické dopravy a přispívá tak k efektivnějšímu plánování budoucí cesty.

Abstract

Shared urban mobility is a continuously evolving phenomenon that, like many others, heavily relies on mobile phones to operate its services. The early stages of development in this sector leave room for improvements, particularly through advanced data analysis, such as the use of predictive models — an approach that remains underutilized in the analysis of shared cycling. These models, however, enable a crucial form of analysis that would allow users to plan their trips based on the likely availability of transport options. The goal of this thesis was to develop a comprehensive analytical tool for shared bicycle transport in the city of Brno, enabling predictions of future availability. A predictive model, trained on historical data from bike-sharing service providers, reliably estimates the probability of a vehicle being present at a selected station at a specific date and time. The result is not only a predictive analysis tool but also a set of insights into the needs and limitations that hinder reliable predictive modeling. This solution offers an innovative approach to analyzing shared cycling data and contributes to more efficient planning of future journeys.

Klíčová slova

cyklistika, bike-sharing, prediktivní analýza, rozhodovací stromy, mobilní aplikace, Brno, .NET

Keywords

cycling, bike-sharing, predictive analysis, decision trees, mobile application, Brno, .NET

Citace

SKŘEČEK, David. *Analýza dat ze služby pro sdílení kol ve městě Brno*. Brno, 2025. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Juraj Lazúr

Analýza dat ze služby pro sdílení kol ve městě Brno

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana Ing. Juraje Lazúra. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal. Při tvorbě této práce jsem využíval nástrojů generativní AI k účelům dohledání užitečné literatury a překladu těchto publikací z anglického jazyka. Rovněž jsem těchto nástrojů užil při implementaci k průzkumu stávajících řešení v rámci tvorby potřebných funkcionalit.

.....

David Skřeček
11. května 2025

Poděkování

Tímto bych rád poděkoval vedoucímu mé práce panu Ing. Juraji Lazúrovi za jeho precizní, avšak přátelský přístup, jeho rady, podněty a podnětnou kritiku. Dík rovněž patří mojí rodině a přátelům za oporu během realizace této práce a během celého předcházejícího studia, stejně jako všem zapojeným do tvorby práce a řešení v rámci ní vytvořeného.

Obsah

1	Úvod	3
2	Nástroje pro analýzu a vizualizaci dat sdílené cyklistiky	4
2.1	Sdílená ekonomika	4
2.1.1	Historie sdílené ekonomiky	4
2.1.2	Oblasti sdílené ekonomiky	5
2.1.3	Sdílená cyklistika	7
2.2	Zdrojová data	8
2.2.1	Serializační formáty	8
2.2.2	Struktura a formát zdrojových dat služby Nextbike	9
2.2.3	Struktura a formát zdrojových dat služby Rekola	10
2.2.4	Získání zdrojových dat	11
2.2.5	Přehled a analýza dostupných endpointů API služby Nextbike	11
2.3	Současné analytické nástroje	12
2.3.1	The Meddin Bike-sharing World Map	12
2.3.2	Bike Share Map	12
2.3.3	Bikesharing Brno	13
2.4	Geografická data	14
2.4.1	Souřadnicový systém	14
2.4.2	Reprezentace a zpracování geografických dat	14
2.4.3	Metody vizualizace	15
2.4.4	Vizualizační nástroje	16
2.5	Shrnutí	18
3	Predikce	19
3.1	Prediktivní modely	19
3.1.1	Klasifikace prediktivních modelů	19
3.1.2	Statistické učení	20
3.1.3	Supervizované a nesupervizované učení	21
3.1.4	Lineární regrese	21
3.1.5	Logistická regrese	21
3.1.6	Užití logistické regrese	22
3.1.7	Rozhodovací stromy	22
3.1.8	Rozhodovací stromy s technikou gradient boosting	22
3.1.9	Užití rozhodovacích stromů	23
3.2	Shrnutí	23
4	Analýza	24

4.1	Uživatelská analýza	24
4.2	Analýza stávajících řešení	26
4.3	Funkční požadavky	26
5	Návrh	27
5.1	Architektura	27
5.1.1	Serverová aplikace a API rozhraní	27
5.1.2	Prediktivní model	28
5.1.3	Mobilní aplikace	29
5.2	Vstupní data	29
5.3	Uživatelské rozhraní	30
6	Implementace	31
6.1	Backend	31
6.1.1	API rozhraní	32
6.1.2	Open Meteo API	33
6.2	Prediktivní model	33
6.2.1	Implementace a trénování	34
6.3	Frontend	35
6.4	Problémy při implementaci	38
6.4.1	Nekonzistence datasetů	38
6.4.2	Mapová knihovna	38
7	Testování	39
7.1	Testování prediktivních modelů	39
7.2	Testování uživatelského rozhraní	41
7.2.1	Uživatel 1	41
7.2.2	Uživatel 2	42
7.3	Závěr testování	42
7.4	Budoucí vývoj	42
8	Závěr	44
	Literatura	45

Kapitola 1

Úvod

Vlivem zhušťující se dopravy ve městech vznikají v městských centrech stále častější dopravní komplikace, které, obzvlášť v dnešní rychlé době, vytvářejí značné překážky lidem ve městech pohybujícím se. Tyto komplikace se často týkají i dopravy hromadné, zejména autobusů, ani tato možnost není vždy řešením. V mnoha ohledech jednou z neadekvátnějších alternativ k dopravě automobilové je cyklistika. V sídlech univerzitních institucí s hojnou částí obyvatel tvořenou studenty, kteří většinou mají pouze přechodné bydliště v onom městě, je provoz osobních nemotorových prostředků nepraktický. Z toho důvodu se stále roste zájem o služby sdílené mobility.

V souvislosti s rostoucími nároky na efektivní a ekologicky šetrnou dopravu se významně zvyšuje i potřeba optimalizace sdílených dopravních služeb, jako je bike-sharing. Tyto služby nejenom že přispívají k udržitelnému rozvoji městské dopravy, ale zároveň odpovídají na konkrétní výzvy městské mobility, kde je kladen důraz na rychlost, dostupnost a minimální dopad na životní prostředí. Tlak ze strany národních i nadnárodních institucí s cílem minimalizovat uhlíkovou stopu navíc využití těchto služeb přímo podporuje. U některých služeb je více než u jiných vyhověno těmto požadavkům, jelikož výroba a likvidace elektrických baterií do elektricky poháněných dopravních prostředků je značně nákladná a ne příliš ekologicky výhodná.

Z toho důvodu lidé volí služby sdílených kol, neboli bike-sharing, jejichž přínos ke snížení produkce emisí je navíc doplněn o podporu zdravého životního stylu, na který je v dnešní době kladen stále větší důraz.

S příchodem těchto služeb je, v rámci optimalizace rozmístění a saturace stanic a kol na stanovištích, nutno data bike-sharingových služeb analyzovat. Z pohledu časové osy je stále nedostatek v analýze vyhlízející do budoucnosti, formou pokročilých analytických metod, jejichž užití pokročilé informační technologie dnešní doby umožňují. V následujících kapitolách bude nastíněn teoretický rámec pro později navrhované a implementované řešení. Kapitola 2 poskytne náhled do základních principů sdílené ekonomiky se zaměřením na sdílenou cyklistiku, geografická data, serializační formáty a možnosti současných analytických nástrojů sdílené cyklistiky. V druhé části, kapitole 3, je popsáno jádro predikce a prediktivních modelů. Dále jsou zde představeny pokročilé analytické metody s rozhodovacími stromy v popředí. Kapitola 4 zkoumá potřeby cílové skupiny, tedy uživatelů služeb sdílené cyklistiky, z jejíž závěrů je vypracován návrh v kapitole 5. Čtvrtá kapitola 6, zaměřující se na implementaci navržené aplikace, podrobně popisuje význačné vlastnosti řešení a překážky, kterým její vývoj čelil. V poslední kapitole 7 je popsáno testování jednotlivých částí řešení a navrženy možnosti pro budoucí vývoj.

Kapitola 2

Nástroje pro analýzu a vizualizaci dat sdílené cyklistiky

Se stále zvyšující se vytižeností služeb sdílené cyklistiky přichází nutnost poptávku analyzovat s cílem optimalizovat a nadále vylepšovat kvalitu nabídky. Pro komplexní pochopení konceptu sdílené cyklistiky a korektní analýzu a její interpretaci je nejprve nutno se seznámit se širším kontextem sdílené ekonomiky, jejíž je sdílená cyklistika součástí.

2.1 Sdílená ekonomika

Jednoznačný pojem sdílené ekonomiky nelze snadno vymezit a ustálená definice doposud neexistuje. V širším slova smyslu sdílenou ekonomiku lze uchopit jako preferování přístupu ke zboží dlouhodobější spotřeby a výrobním faktorům za pomoci digitálních platforem spíše než jejich samotné vlastnictví. Tímto jsou snižovány transakční náklady, což přímo přispívá ke snížení nájemní ceny takových sdílených zdrojů a zvyšuje se tak dostupnost širšímu okruhu uživatelů [24]. Podle odhadu MIT je běžná využitelnost osobních automobilů v USA pouhých 5% [15]. Analýzy služeb, jako je Car2Go, ukazují, že každý sdílený automobil může přitom odstranit z městských silnic až 11 soukromých vozů, což vede ke snížení skleníkových plynů až o 10% [14].

2.1.1 Historie sdílené ekonomiky

Fenomén sdílené ekonomiky s sebou nese letitou historii, sahající do dob komunitního sdílení a barterového obchodu. S rozvojem digitálních technologií a internetu přišel rozmach sdílené ekonomiky díky dostupnosti, kterou technologický pokrok zpřístupnil. Dnešní podoba sdílené ekonomiky, mezi jejíž hlavní protagonisty patří platformy jako Airbnb, Uber nebo Nextbike, je přímo ovlivněna přirozeně vyvíjejícími se právy na osobní vlastnictví. Její počátky však sahají až do dob pravěku, kdy se objevily prvobytné elementy v podobě komunitního sdílení.

- **Tradiční komunitní sdílení**

Sdílená ekonomika začala v podobě komunitního sdílení a barterového obchodu, kdy lidé v rámci malé komunity sdíleli nástroje, zemědělské prostředky a pracovní sílu. Tento model fungoval na základě vzájemné důvěry a reciprocity. Historické příklady zahrnují společné užívání zemědělské půdy nebo systém výměnného obchodu, známý jako barter [2].

- **Počátky digitálního sdílení**

S rozvojem internetu a webových stránek na konci 20. století se objevily první pokusy o digitální sdílení. Platformy jako eBay (založeno 1995) umožnily jednotlivcům přímo prodávat a nakupovat použité zboží. [8] Tato fáze představuje základní kámen pro rozvoj online tržišť a sdílené ekonomiky, která podporuje individuální podnikání a snižuje závislost na tradičních maloobchodních řetězcích [22].

- **Vzestup sdílené ekonomiky**

S rychlým rozvojem technologií v letech 2000–2010 vznikají nové platformy zaměřené na sdílení, jako jsou Airbnb a Couchsurfing. Tyto služby umožnily uživatelům sdílet své domovy nebo nevyužívané nemovitosti s ostatními za poplatek. Sdílení dopravy získalo popularitu díky vzniku Uberu (2009), který revolucionalizoval dopravu ve městech tím, že umožnil uživatelům snadno a rychle najít dostupné vozidlo s užitím mobilních aplikací. Tento model nejen usnadnil přístup k dopravě, ale také přispěl k omezení počtu soukromých vozidel ve městech a zlepšil efektivitu dopravy, čímž přispěl ke snížené produkci emisí [2].

- **Regulace a normalizace**

Okolo roku 2010 začaly vlády a instituce řešit právní a ekonomické důsledky sdílené ekonomiky. Platformy jako Uber a Airbnb se staly předmětem sporů týkajících se pracovních podmínek, daňové legislativy a regulace. Toto období přináší i první pokusy o nastavení pravidel a zajištění ochrany práv jak pracovníků, tak spotřebitelů [8].

- **Současná podoba sdílené ekonomiky**

Současný trend směřuje k využití decentralizovaných technologií, jako je Blockchain, které podporují ještě větší transparentnost a samoregulaci. Tyto technologie umožňují peer-to-peer transakce bez potřeby centrálního zprostředkovatele, což snižuje náklady a zvyšuje důvěryhodnost. Příkladem může být vznik decentralizovaných tržišť nebo sdílené sítě energie, jako jsou komunitní solární projekty [22].

2.1.2 Oblasti sdílené ekonomiky

Koncept sdílené ekonomiky přináší do mnoha oblastí inovativní způsoby využití zdrojů, napomáhající tím snížení nákladů. Mezi klíčové sektory patří služby finanční, ubytovací a doprava, kde dochází k rozvoji platform propojujících poskytovatele a poptávající.

- **Finanční služby**

Sdílená ekonomika se také začíná prosazovat jako alternativní forma financování. Příkladem je crowdfunding, neboli peer-to-peer funding, kdy se na úvěr jednomu dlužníkovi či na jeden projekt složí více investorů. Každý z investorů přitom zpravidla půjčuje subjektu relativně malou částku, čímž diverzifikuje riziko. Crowdfundingové platformy většinou pouze zprostředkovávají přesun finančních prostředků mezi investory a dlužníky, nejsou tedy přímo poskytovateli úvěrů. Úvěry mohou být poskytovány jak spotřebitelům, tak začínajícím firmám. Příklady firem působících v oblasti finančních služeb jsou Kickstarter 2.1, GoFundMe a Indiegogo [4].

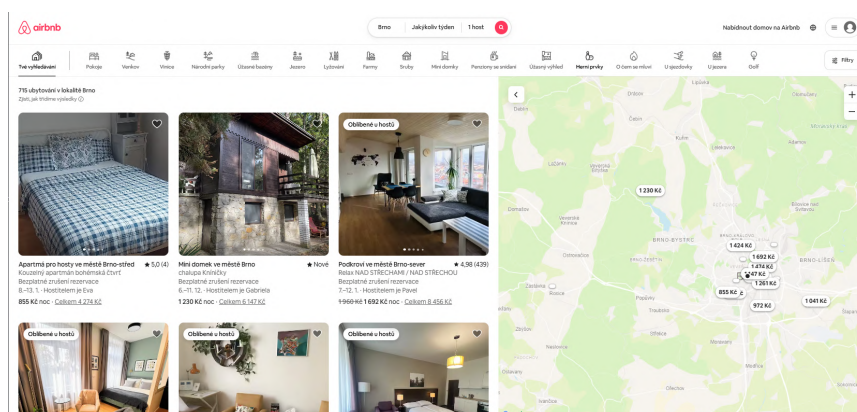
Společnosti Kickstarter má za cíl přivést kreativní projekty k životu prostřednictvím crowdfundingu. Jedním z nich je například projekt Eclipta 2.1, zabývající se výrobou solárních panelů zabudovaných do skel oken.



Obrázek 2.1: Profil projektu Eclipta na platformě Kickstarter

• Ubytování

Dalším aktivem s potenciálem pro využití v rámci sdílené ekonomiky jsou obytné prostory. Nejrozšířenější platformou umožňující sdílet krátkodobě nemovitost je Airbnb 2.2. Tyto služby jsou atraktivní jak pro vlastníky nemovitostí z hlediska celkového vyššího výnosu z krátkodobých pronájmů, tak pro uživatele, pro které je cena těchto služeb často přívětivější než ceny služeb hotelů [23].



Obrázek 2.2: Nabídka ubytování na platformě Airbnb ve městě Brno

• Doprava

Za cílem snížení dopadu emisí, které přispívají ke klimatickým změnám, a také důsledkem vysoké pořizovací a provozové ceny automobilů se stále více rozrůstá trend sdílené dopravy. Značným způsobem pak přispívá k dopravní situaci ve městech. Řada majitelů aut jezdí do práce i z práce sama, auta jsou poloprázdná a často stojí dlouhé dny nevyužitá v garáži. Jedno sdílené auto má potenciál nahradit téměř deset soukromých vozidel [17]. Kromě sdílených automobilů představuje ekologicky i ekonomicky výhodnou alternativu osobní automobily sdílená cyklistika. Mezi nejsilnější platformy na českém trhu patří Bolt, který se kromě klasického ride-hailingu nově zaměřuje i na car-sharing, a Nextbike, lídr v oblasti sdílené cyklistiky, který působí ve více než 40 českých městech. Dalšími významnými platformami je

Liftago, známé především pro své služby ride-hailing, které dávají uživatelům možnost objednat si jízdu a zvolit si řidiče podle svých preferencí, a společnost Rekola [2.3](#), která je předním českým zástupcem v oblasti sdílené cyklistiky.



Obrázek 2.3: Ilustrační foto společnosti Rekola

2.1.3 Sdílená cyklistika

Sdílená cyklistika, známá také jako bike-sharing, zahrnuje sdílení jízdních kol pro krátkodobé využití. Systémy tohoto typu nabízejí nejen klasická kola, ale také elektrokola a elektrické koloběžky, které se v posledním desetiletí rozšířily i do menších měst díky podpoře městských správ, které tak chtějí zvýšit kvalitu života a usnadnit ekologickou dopravu pro své obyvatele [\[20\]](#).

Vývoj sdílené cyklistiky

První systémy sdílení kol vznikly v Evropě v 60. letech, kdy byly veřejnosti bez dohledu zpřístupněny jednoduché bicykly. Moderní podoba sdílené cyklistiky se však rozšířila až v 90. letech s růstem městských investic do udržitelné dopravy. Značným milníkem byl program Vélib' spuštěný v Paříži v roce 2007, který inspiroval mnoho dalších měst k vytvoření vlastních systémů sdílení kol [\[5\]](#).

Moderní bike-sharingové systémy stále častěji zahrnují nejen klasická kola, ale i elektrokola a elektrické koloběžky, což rozšiřuje jejich dosah a umožňuje uživatelům pokrývat větší vzdálenosti. Díky propojení s mobilními aplikacemi mohou uživatelé snadno najít a rezervovat volná kola, což přináší vyšší pohodlí a snadnější dostupnost [\[20\]](#).

Přestože popularita sdílené cyklistiky neustále roste, mnoho měst se stále potýká s problémy, jako je údržba kol, vandalismus a nepoměr mezi nabídkou a poptávkou během exponovaných časů. V budoucnu se očekává, že využití technologií, jako jsou big data a umělá inteligence, umožní lépe předpovídat poptávku a optimalizovat rozmístění kol po městě, čímž se zlepší celková efektivita těchto systémů [\[6\]](#).

Nejen v České republice však v posledních letech značně přibyl počet dopravních nehod spojených s provozem elektrokoloběžek, což vede ke zpřísnění bezpečnostních opatření u nás [\[3\]](#). K úplným zákazům pak nehodovost vedla primárně ve větších městech, jako je Paříž, kde úplný zákaz sdílených elektrokoloběžek platí již od roku 2023 [\[7\]](#).

2.2 Zdrojová data

Pro vhodnou interpretaci dat a jejich analýzu je nejprve nutné porozumět jejich formátu, způsobu získání a vyhodnotit jejich kvalitu. Tato data mohou být obstarána prostřednictvím oficiálních rozhraní daných služeb, pomocí automatizovaných systémů (např. GPS), či mohou být výsledky modelů a simulací.

Tato kapitola se podrobně zaměřuje na podobu a cestu k obstarání dat z konkrétních služeb působících ve městě Brno a na obecné vlastnosti formátů, které jsou využívány pro přenos a uchování těchto dat.

2.2.1 Serializační formáty

V informatice jsou serializační formáty nedílnou součástí pro přenos a ukládání strukturovaných dat. Umožňují převést různé datové struktury, mnohdy velice složité, na formát, který je lehce čitelný a zpracovatelný. Mezi nejpoužívanější formáty patří XML a JSON, často využívané ve webových aplikacích, zpřístupňující komunikaci mezi servery a klienty.

Serializační formáty dat jsou klíčové pro efektivní přenos a ukládání strukturovaných informací. Mezi nejpoužívanější formáty patří XML a JSON:

1. XML

XML, neboli Extensible Markup Language, je flexibilní, formátově jednoduchý značkovací jazyk hojně používaný pro svou snadnou přenositelnost a rozšiřitelnost. Značky, pomocí kterých se data zapisují, jsou plně modifikovatelné, nezávislé na jakékoliv specifické platformě či software, uživatel má možnost si definovat značky a struktury vlastní.

Příklad struktury XML

```
<countries>
  <country name="Czech Republic" country="cz">
    <city name="Brno" lat="49.1951" lng="16.6068" />
  </country>
</countries>
```

2. JSON

JSON, neboli JavaScript Object Notation, je podobně jako XML flexibilní a formátově jednoduchý. Původně vyvinut pro přenos dat zejména v prostředí JavaScriptu, je dnes podporovaný celou řadou programovacích jazyků, díky tomuto, a své efektivitě, je velice populární pro přenos dat.

Příklad struktury JSON

```
{
  "countries": [
    {
      "name": "Czech Republic",
      "countryCode": "cz",
      "cities": [
        {
          "name": "Brno",
          "lat": 49.1951,
          "lng": 16.6068
        }
      ]
    }
  ]
}
```

2.2.2 Struktura a formát zdrojových dat služby Nextbike

Služba Nextbike poskytuje data prostřednictvím svého API ve formátech XML a JSON. API umožňuje získání aktuálních informací o stanicích, kolech a jejich dostupnosti, aktuálních výpůjčkách, atp. ve všech městech, včetně Brna.

Hlavní prvky datového modelu

API poskytuje několik klíčových prvků, které tvoří strukturu dat:

- **countries:** Kořenový element, který obsahuje všechny země, kde Nextbike své služby provozuje.
- **country:** Obsahuje informace o konkrétní zemi jako **name** (název země), **country** (kód země) a **lat/lng** (souřadnice středu země).
- **city:** Obsahuje atributy o konkrétním městě, např. **name** (název města), **lat/lng** (souřadnice středu města) a **num_places** (počet stanic ve městě).
- **place:** Tento prvek představuje jednotlivé stanice. Obsahuje atributy jako **name** (název stanice), **lat/lng** (souřadnice stanice), **bikes** (počet dostupných kol) a **bike_types** (typy dostupných kol).
- **bikes:** Počet prostředků nacházejících se na stanici.
- **bike_list:** Seznam kol doplňujících informace o jednotlivých prostředcích, jako jsou **number** (identifikátor provozovatele), **state** (provozní schopnost prostředku), aj.

Příklad struktury XML dat

```
<?xml version="1.0" encoding="UTF-8"?>
<countries>
  <country name="Czech Republic" country="cz" lat="49.8175" lng="15.4730">
    <city uid="660" alias="Nextbike Brno" lat="49.1951" lng="16.6068">
      <place uid="27618846" lat="49.194588" lng="16.608399"
        name="Náměstí Svobody JZ" bikes="6">
        <bike number="481830" active="1" state="ok" />
        <bike number="481756" active="1" state="ok" />
        <bike number="486886" active="1" state="ok" />
      </place>
    </city>
  </country>
</countries>
```

Klíčové atributy

1. **name**: Název země, města nebo stanice.
2. **lat** a **lng**: Souřadnice jednotlivých stanic.
3. **bikes**: Počet dostupných prostředků na stanici.
4. **number**: Unikátní identifikátor kola.

Tato struktura umožňuje detailní analýzu rozmístění stanic, dostupnosti kol a dalších parametrů služby Nextbike v Brně. Data poskytovaná tímto API jsou vhodná pro dynamické sledování a optimalizaci využívání sdílených kol ve městě.

2.2.3 Struktura a formát zdrojových dat služby Rekola

Narozdíl od služby Nextbike neposkytuje provozovatel Rekola veřejně dostupné API pro získání dat. Korepondenci lze však obdržet záznamy o jízdách ve formě .csv souborů.

Hlavní prvky datového modelu

Služba Rekola poskytuje ve svých surových datech několik klíčových prvků, které tvoří jejich strukturu:

- **start_time** / **end_time**: Datum a čas začátku a konce jízdy.
- **timediff_seconds**: Délka trvání jízdy v sekundách.
- **tariff**: Tarifní zóna.
- **user_id**: Identifikátor uživatele, který jízdu uskutečnil.
- **start_rack_name** / **end_rack_name**: Název počáteční a cílové stanice.
- **start_rack_latitude** / **start_rack_longitude**: Souřadnice počáteční stanice.
- **end_rack_latitude** / **end_rack_longitude**: Souřadnice cílové stanice.

- **start_rack_id / end_rack_id**: Identifikátor počáteční a cílové stanice.
- **registration_datetime**: Datum a čas registrace uživatele.
- **loan_id**: Identifikátor půjčky spojené s jízdou.

2.2.4 Získání zdrojových dat

K získání dat pro analýzu služby Nextbike slouží veřejné API této společnosti, jež disponuje sadou endpointů poskytujících určité informace o službě, jako například aktuální obsazenost stanic, stav jednotlivých kol apod.

Získání dat služby Rekola není možné provádět automatizovaně, neexistuje veřejně dostupné API. Z existujících záznamů v podobě tabulek formátu .csv sice nelze data analyzovat v reálném čase, lze ale vytvářet analýzy za uplynulá kalendářní období.

2.2.5 Přehled a analýza dostupných endpointů API služby Nextbike

API služby Nextbike poskytuje řadu endpointů, které lze využít k získání různých typů dat potřebných pro analýzu provozu a využití sdílených kol ve městě Brno. V této sekci jsou jednotlivé endpointy popsány z hlediska jejich funkcionality a relevance pro analytické účely.

/maps

Endpoint **/maps** je jedním z nejdůležitějších pro základní analýzu, protože poskytuje aktuální informace o všech dostupných stanicích a jejich obsahu (počet kol a volných míst) v reálném čase. Tento endpoint je užitečný pro:

- Získání stavů všech stanic v určitém městě, včetně Brna.
- Získání dat o konkrétní stanici.

Data získaná z tohoto endpointu jsou klíčová pro analýzu rozložení kol po městě a identifikaci vytížených stanic.

/rentals

Endpoint **/rentals** slouží ke správě uživatelů a výpůjček. Tento endpoint by mohl být zejména užitečný aplikacím sjednocujícím funkce konkurenčních služeb.

Shrnutí

Pro účely analýzy dat služby Nextbike ve městě Brno by mohl být nejpoužitelnějším zejména endpoint **/maps**. Tento endpoint poskytuje kritické informace o aktuálním stavu stanic a dostupnosti kol, což je nezbytné pro detailní analýzu využívání služby a optimalizaci jejího provozu.

Data z tohoto endpointu mohou být zpracována a analyzována k identifikaci trendů v chování uživatelů, optimalizaci umístění stanic a zajištění dostatečné dostupnosti kol v různých částech města.

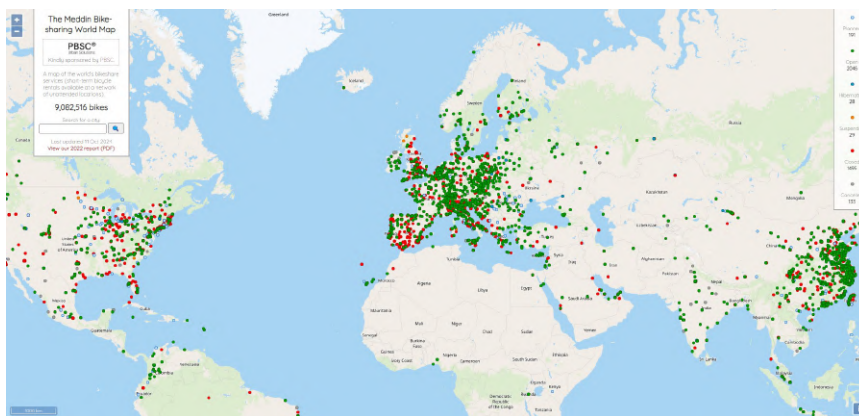
2.3 Současné analytické nástroje

Se vzestupem sdílené cyklistiky se zrodil problém služby analyzovat pro zlepšení provozuschopnosti. Kromě nástrojů provozovatelů existuje řada volně dostupných, často kombinujících data několika společností, tvořících tak přehlednější, ale obsahově méně detailní analýzu.

2.3.1 The Meddin Bike-sharing World Map

Interaktivní mapa The Meddin Bike-sharing World Map ¹ nabízí přehled služeb sdílené cyklistické dopravy aktuálně provozovaných, také však projektů zrušených anebo plánovaných. Uživatel má možnost zobrazení detailu dané služby v daném městě, takto může přistoupit k datům o počátku, případně o ukončení provozu, v závislosti na dostupnosti interních dat provozovatelů, pak také k jednoduchým analýzám vytíženosti v závislosti na čase.

Aplikace uživatelům nabízí globální přehled zahrnující informace o stavu provozu a případných jiných metrikách. Co naopak uživatel v aplikaci nenajde, jsou údaje o konkrétních stanicích a analytické údaje v rámci nich ^{2.4}.



Obrázek 2.4: Úvodní obrazovka aplikace The Meddin Bike-sharing World Map

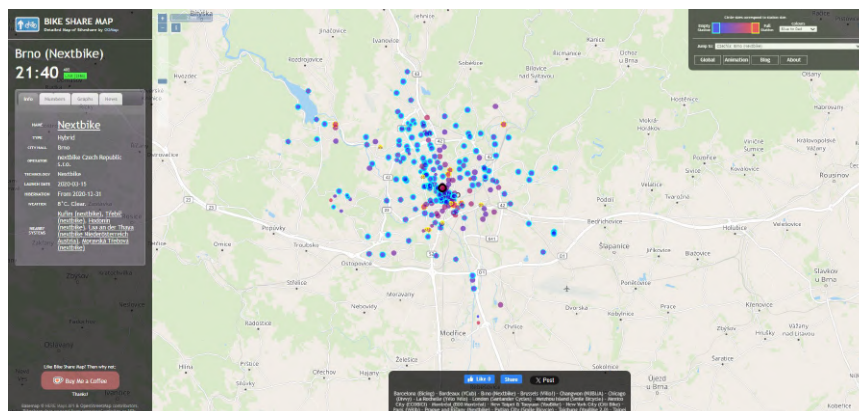
2.3.2 Bike Share Map

Aplikace Bike Share Map ² řeší problém The Meddin Bike-sharing World Map, přichází kromě globálního přehledu služeb ve městech také s možností omezení se na pouze vybranou oblast a konkrétní stanice, které se v ní nacházejí. Ty jsou také doplněny o analytická data vytíženosti daného uzlu z pohledu výpůjček a obsazenosti míst.

Uživatel má přístup k detailnější analýze dat poskytovatelů služeb s možností omezit se pouze na vybrané stanice. V mnoha městech, včetně Brna, je ale nástroj omezen pouze na jednoho či hrstku provozovatelů. Nekompletní je pak i seznam stanic na mapu zanesených ^{2.5}.

¹<https://bike-sharingworldmap.com/>

²<https://bikesharemap.com/>



Obrázek 2.5: Bike Share map s vizualizací stanic v Brně

2.3.3 Bikessharing Brno

Aplikace Bikessharing Brno³, jejíž autory jsou Iva Benešová, Daniela Jurečková a Ondřej Doseděl, se zaměřuje, jak název napovídá, na analýzu sdílené cyklistiky ve městě Brně, a to způsobem podstatně podrobnějším, než již zmíněná konkurence. Namísto dat klasických, jako jsou lokální průměry délky jízd a sumy výpůjček, nabízí také doplňkové informace, jako je teplotní analýza v rámci jízd aj.

Řešení výrazně detailněji řeší datovou analýzu poskytovatelů sdílené cyklistické dopravy. Napříč tomu je ale aplikace zaměřena výhradně na město Brno a pod jeho správu spadající blízké okolí 2.6.



Obrázek 2.6: Teplotní analýza aplikace Bikessharing Brno

³<https://bikessharingbrno.streamlit.app/>

2.4 Geografická data

V době chytrých telefonů je prakticky nemožné opomenout vytíženost mapových aplikací. Nejen pro cestování jsou tyto aplikace vybaveny, obsahují nespočet informací o podnicích všeho druhu, benzínových stanicích, historických památkách atp., které často bývají doplněny ikonami. Pro zobrazení v mapě je třeba zvolit vhodný souřadnicový systém, pomocí něhož bude možné jednotlivé body jednoznačně zanechat do prostoru.

2.4.1 Souřadnicový systém

Jednoznačné identifikování bodů a jiných útvarů v prostoru je možné pomocí zeměpisných souřadnic. Poloha každého bodu je popsána zeměpisnou šířkou a délkou. Tyto číselné hodnoty vyjadřují úhel, který svírá přímka vedoucí od bodu ke středu Země s rovinou rovníku pro zeměpisnou šířku a s rovinou nultého poledníku pro zeměpisnou délku.

Veškeré body s totožnou zeměpisnou šířkou tvoří rovnoběžky, neboli paralely, které jsou kružnicemi rovnoběžnými s rovníkem. Body se stejnou zeměpisnou délkou tvoří poledníky, jinak také meridiány, které se táhnou mezi póly. Zeměpisné souřadnice jsou určeny dvěma hlavními hodnotami: zeměpisnou šířkou a zeměpisnou délkou, které se definují takto:

- Zeměpisná šířka nabývá hodnot 0° až 90° severní nebo jižní, kde 0° představuje rovník.
- Zeměpisná délka se vyjadřuje hodnotami 0° až 180° východní nebo západní délky, přičemž 0° označuje nultý poledník.

Pro práci s geografickými daty je důležité správně zvolit souřadnicový systém. Například WGS-84 (World Geodetic System) je široce používaným systémem, zatímco místní systémy, jako S-JTSK v České republice, jsou přesnější pro konkrétní území [11].

2.4.2 Reprezentace a zpracování geografických dat

Geografická data jsou v geoinformačních systémech reprezentována standardizovanými formáty, díky kterým je zobrazení konzistentní a efektivní. Mezi nejčastěji používané formáty patří [11]:

- **Vektorová data**

Vektorová data jsou založena na geometrických tvarech – bodech, liniích a polygonech. Jsou ideální pro reprezentaci objektů, jako jsou stanice sdílených kol (body), trasy cyklostezek (linie) nebo oblasti pokrytí službami (polygony). Každý prvek může obsahovat atributy popisující jeho vlastnosti (např. název stanice, kapacita kol) [13].

- **Rastrová data**

Rastrová data využívají mřížku pixelů pro reprezentaci spojitých geografických jevů, například satelitních snímků, map teplot nebo heatmap využití sdílených kol. Každý pixel obsahuje hodnotu reprezentující intenzitu jevu (např. počet vypůjčení kol) [11].

Zpracování těchto dat obvykle zahrnuje následující kroky:

- **Transformace souřadnic**

Převod dat mezi různými souřadnicovými systémy (např. z WGS-84 do S-JTSK) zajišťuje přesnou reprezentaci polohy objektů.

- **Úprava a analýza dat**

Zahrnuje filtrování, shlukování a analýzu atributů, a prostorovou analýzu vztahů mezi geografickými objekty.

- **Interpolace**

Vytváření spojitých ploch z bodů, například generování teplotních map z naměřených hodnot.

2.4.3 Metody vizualizace

Úloha vizualizace v geografických analýzách se netýká jen map a údajů z DPZ či fotogrammetrie, ale stejně tak se týká např. statistických analýz. Interpretace grafu nebo schématu je efektivnější než interpretace založená na množství čísel, reprezentujících tytéž údaje [11].

Metody vizualizace geografických dat zahrnují různé přístupy, které umožňují efektivní interpretaci prostorových informací. Cílem vizualizace je nejen prezentovat data, ale také zprostředkovat hlubší pochopení vzorců a trendů, jež nejsou na první pohled patrné. Lze je kategorizovat do následujících skupin:

- **Kartografické metody**

Zahrnují tradiční i moderní způsoby zobrazení geografických dat pomocí map. Tematické mapy, jako například choropleth mapy, zobrazují statistická data za použití techniky barevné symbolizace (např. hustotu využití sdílených kol v jednotlivých městských částech). Dalším příkladem jsou izoliniové mapy, které využívají spojitě křivky pro znázornění hodnot stejné intenzity (např. nadmořské výšky či koncentrace znečištění). Topografické mapy kombinují fyzické vlastnosti krajiny, jako jsou terén a vodní plochy, s informacemi o infrastruktuře a zástavbě [11].

- **Grafické metody**

Grafické metody využívají prvky jako body, čáry, symboly či grafy pro vizuální prezentaci dat. Příkladem může být bodový graf, který ukazuje rozmístění jednotlivých stanic sdílených kol v prostoru, nebo sloupcový graf, který srovnává četnost jejich využití v různých časových intervalech. Dalšími možnostmi jsou teplotní mapy (heatmapy), které ukazují hustotu jevů v dané oblasti pomocí barevného gradientu [13].

- **3D vizualizace**

Třetí rozměr umožňuje detailnější prezentaci krajiny nebo urbanistických struktur. Například 3D modely měst mohou vizualizovat propojení cyklostezek a jejich výškové profily. Tento způsob vizualizace se často používá pro prezentaci dat získaných z lidarových měření nebo digitálních modelů terénu (DTM). Užitečné je i zobrazení trajektorií pohybu v trojrozměrném prostoru, což je relevantní například při analýze pohybu sdílených kol ve velkoměstech [11].

- **Animace a dynamické vizualizace**

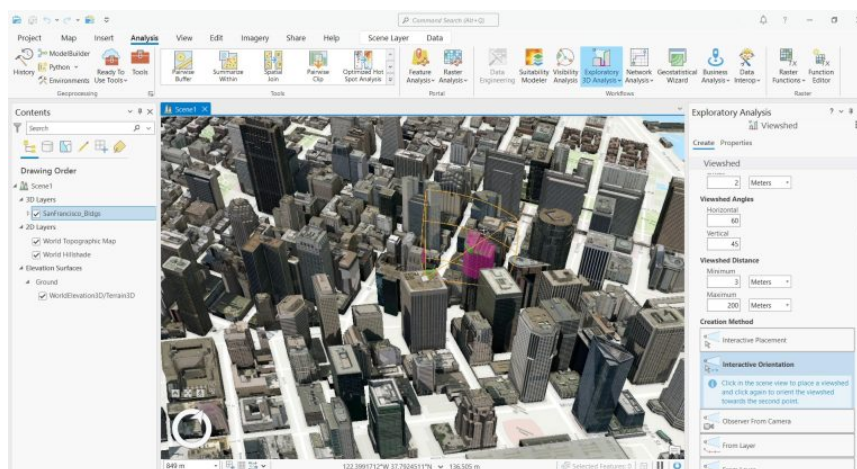
Dynamické metody umožňují sledovat změny geografických dat v čase. Typickým příkladem je časosběrná vizualizace ukazující, jak se mění počet zapůjčených kol během dne nebo jak se vyvíjí hustota dopravy v průběhu týdne. Animace přidávají datům dimenzi času a usnadňují porozumění časoprostorovým trendům, které jsou obtížně zachytitelné statickými metodami [16].

2.4.4 Vizualizační nástroje

Geografická data jsou často pro člověka nečitelná či nic neříkající. K jejich interpretaci člověkem nejen v oboru informatiky slouží různé vizualizační nástroje, podle jejichž základních vlastností můžeme tyto nástroje rozdělit do několika kategorií. Mezi hlavní patří:

- **Geoinformační systémy**

Specializované informační systémy, jejichž primárním cílem je sbírat, ukládat, analyzovat a vizualizovat geografická data. Tato data obsahují informace o poloze, tedy kde se daný objekt nebo jev nachází, o jeho obsahu či vlastnostech, například konkrétní událost nebo popisovaný jev, a v některých případech také o čase, kdy k danému jevu došlo nebo kdy byl zaznamenán [13]. Hojně používaným geoinformačním systémem je ArcGIS 2.7.



Obrázek 2.7: Geografický informační systém ArcGIS

- **Statistické nástroje**

Statistické softwary umožňují analýzu a vizualizaci dat s důrazem na kvantitativní interpretaci. Příklady zahrnují:

- **R** – Statistické výpočetní prostředí s knihovnami jako `ggplot2` a `sf`, které podporují tvorbu grafů a map.
- **Python** – Programovací jazyk s knihovnami jako `Matplotlib`, `Geopandas` a `Folium` pro prostorovou analýzu a interaktivní vizualizace.

Statistické nástroje, jakým je například `Matplotlib` 2.8, jsou ideální pro vytváření analytických dashboardů kombinujících grafy, tabulky a mapy.

2.5 Shrnutí

Sdílená cyklistika představuje důležitou část sdílené ekonomiky a řeší výzvy efektivní dopravy ve městech i mimo ně. Správná interpretace a vizualizace těchto dat je klíčová pro analýzu služeb sdílené cyklistiky a pro vývoj optimálních a snadno použitelných analytických nástrojů, které poskytují relevantní informace.

Výběr vhodných serializačních formátů, potřebných pro přenos a ukládání dat, na kterých probíhá konkrétní analýza, je nezbytný pro efektivní a flexibilní práci s daty. Mezi přední představitele těchto formátů patří XML a JSON pro obecná data, nebo jejich specializované varianty pro geografická data, jako jsou GML a GeoJSON.

Geografické informační systémy (GIS) a další nástroje pro vizualizaci geografických dat poskytují interaktivní analytické možnosti, které usnadňují interpretaci prostorových dat a podporují rozhodování v reálném čase. Tato interaktivita je klíčová pro optimální využívání dat o sdílené cyklistice, zejména při analýze uživatelských vzorců nebo plánování rozvoje těchto systémů.

Kapitola 3

Predikce

Oblast učení a následné předpovědi budoucích událostí, neboli predikce, je jednou z možností, které moderní počítači poháněná statistika přináší. Učit se rozpoznávat vzory a na jejich základě predikovat budoucí události je úloha prediktivních modelů [9].

3.1 Prediktivní modely

Statistické nebo strojově učené algoritmy určující budoucí výsledky v závislosti na historických datech, neboli prediktivní modely, jsou mechanismy hojně používané v mnoha oblastech. Mnoho přístupů používaných v prediktivních modelech je založeno na principech statistického učení. Od personalizovaného složení medikamentů po plánování tras, prediktivní modely jsou nedílnou součástí moderních praktik v mnoha odvětvích [9].

3.1.1 Klasifikace prediktivních modelů

Prediktivní modely lze rozdělit do několika kategorií v závislosti na několika klíčových metrikách [9]. Rozlišením různých skupin modelů je umožněno přizpůsobit způsob predikce specifickým požadavkům dané úlohy, jako je typ dat, požadovaný výstup, složitost výpočtu nebo požadavky na přesnost.

- **Podle typu výstupu:**

- **Klasifikační modely**

Modely, jejichž výstupem je přiřazení dat do jedné z několika kategorií. Tyto modely používají algoritmy jako rozhodovací stromy, logistickou regresi nebo neuronové sítě. Příkladem může být úloha rozpoznávání spamu v e-mailové poště, kde cílem je rozhodnout, zda je zpráva spam, nebo ne.

- **Regresní modely**

Modely, které poskytují spojitou hodnotu jako výstup. Regresní modely používají techniky jako lineární regrese, Ridge regrese nebo hluboké neuronové sítě. Příkladem může být predikce cen nemovitostí na základě různých atributů.

- **Podle použité metody:**

- **Statistické modely**

Tyto modely se zakládají na tradičních statistických technikách, jako jsou například lineární modely nebo metoda nejmenších čtverců. Jsou vhodné pro analýzu vztahů mezi proměnnými a predikci výsledků na základě těchto vztahů. Příkladem je použití logistické regrese pro modelování pravděpodobnosti určitého výstupu.

- **Strojové učení**

Modely strojového učení, jak je popsáno v knize *The Elements of Statistical Learning* [9], zahrnují algoritmy, které se učí ze vzorců v datech, aby poskytly predikce nebo rozhodnutí na základě historických dat. Příkladem je použití algoritmu Support Vector Machines (SVM) nebo rozhodovacích stromů.

- **Pravděpodobnostní modely**

Tato metoda vychází z teorie pravděpodobnosti a modeluje vztahy mezi vstupy a výstupy na základě pravděpodobnostních distribucí. Tyto modely zahrnují například Bayesovské sítě nebo modely jako Naive Bayes, které se často používají při klasifikaci textů nebo jiných datových struktur.

- **Podle způsobu učení:**

- **Supervizované učení**

V tomto případě je model trénován na historických datech, kde jsou známé vstupy i odpovídající výstupy. Cílem je naučit model tak, aby na základě nových, neoznačených dat byl schopen správně předpovědět výstupy. Příkladem je trénování modelu na historických datech o počasí k predikci teploty v budoucnosti [9].

- **Nesupervizované učení**

Učení bez označených dat, kde model hledá strukturu nebo vzory v samotných datech. Tento typ učení je často používán pro shlukování nebo dimenzionální redukci. Příkladem může být analýza segmentace zákazníků na základě jejich chování bez předem definovaných skupin [9].

Supervizované a nesupervizované učení představují speciální formy statistických metod, které se liší informacemi tvořícími vstupní data a cíli analýzy.

3.1.2 Statistické učení

Statistické učení je oblast, která se zabývá metodami pro analýzu a interpretaci složitých datových souborů. Cílem statistického učení je identifikovat vzorce a vztahy v datech, aby bylo možné předpovídat budoucí výsledky nebo lépe porozumět procesům, které data generují. Jedná se o širokou disciplínu, která zahrnuje jak tradiční statistické metody, tak moderní přístupy, jako je strojové učení a data mining. Statistické učení je klíčové pro aplikace v oblastech, jako jsou finance, biomedicína, marketing a mnoho dalších. Ačkoliv se statistické učení dělí na různé techniky a přístupy, jeden z hlavních rozdílů spočívá v tom, zda máme k dispozici výstupy, které model učíme předpovídat (supervizované učení), nebo pouze hledáme skryté struktury ve vstupních datech (nesupervizované učení) [10].

3.1.3 Supervizované a nesupervizované učení

Supervizované učení zahrnuje práci s daty, kde máme vstupy (nezávislé proměnné) a odpovídající výstupy (závislá proměnná). Cílem je vytrénovat model, který dokáže předpovídat výstup pro nové vstupy. Příklady zahrnují regresi a klasifikaci, např. posuzování, zda je emailová komunikace spam, či nikoliv [9]. Kupříkladu společnost Google využívá kombinaci strojového učení a uživatelské zpětné vazby, čímž je zajištěno zachycení až 99,9% případů spamu a phishingu [12].

Nesupervizované učení se zabývá daty, která obsahují pouze vstupy bez známých odpovídajících výstupů. Cílem je odhalit skryté struktury nebo vzory v datech. Mezi typické metody patří shlukování (clustering) a dimenzionální redukce [10], pomocí kterých je možné například seskupovat uživatele do skupin dle podobného chování nebo detekovat anomálie. Knihovna Anomalib je jedním z praktických uplatnění a slouží k detekci anomálií v průmyslových datech, například při kontrole kvality výrobku [1].

3.1.4 Lineární regrese

Pro předpověď budoucích událostí, jakou je například pravděpodobný konkrétní počet prostředků na stanici, se používají modely lineární regrese. Lineární regrese je statistická metoda používaná k modelování vztahu mezi závislou proměnnou Y a jednou nebo více nezávislými proměnnými $X_1, \dots, X_p$, kde $p \geq 1$. Cílem je nalézt lineární funkci, která co nejlépe přizpůsobí data. Lineární regresi lze použít k predikci hodnoty závislé proměnné na základě hodnot nezávislých proměnných. Příkladem může být predikce ceny nemovitosti na základě její velikosti a lokality, nebo predikce výšky člověka na základě věku a pohlaví [9].

3.1.5 Logistická regrese

Logistická regrese je speciální formou lineární regrese, která je určena pro klasifikaci, kdy výstupem modelu je pravděpodobnost, že dojde k určité události. Stejně jako u lineární regrese, model využívá lineární kombinaci vstupních proměnných $X_1, \dots, X_p$, avšak výstup této funkce je aplikován na logistickou funkci, která omezuje výsledek na interval $[0, 1]$, což umožňuje interpretovat výstup jako pravděpodobnost. Model logistické regrese je definován následovně [9]:

$$P(Y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}$$

kde:

- $P(Y = 1 \mid \mathbf{x})$ je pravděpodobnost, že výstupní proměnná Y nabývá hodnoty 1, podmíněně na vektor vstupních proměnných $\mathbf{x}$.
- $\mathbf{x} = (x_1, x_2, \dots, x_n)$ je vektor vstupních proměnných.
- $\beta_0, \beta_1, \dots, \beta_n$ jsou parametry modelu, které se naučí během trénování.

3.1.6 Užití logistické regrese

Logistická regrese nabývá širokého využití v mnoha oblastech díky své schopnosti předpovídat vztahy mezi nezávislými proměnnými a pravděpodobností výskytu binární závislé proměnné. Je možné se s ní setkat v následujících odvětvích [19]:

- **Zdravotnictví**
Předpověď výskytu onemocnění na základě rizikových faktorů.
- **Finance**
Aplikace logistické regrese ve strojovém učení pro odhalení finančních podvodů.
- **Marketing**
Modelování chování zákazníka v kontextu užívání produktu.

3.1.7 Rozhodovací stromy

Rozhodovací stromy jsou nelineární modely, které rozdělují prostor vstupních vlastností do oblastí, v rámci kterých přizpůsobují jednoduchý model. Na rozdíl od lineárních metod, jako jsou lineární a logistická regrese, které předpokládají lineární vztah mezi vlastnostmi a výsledkem, rozhodovací stromy předpokládají vztah nelineární, tedy vztah, který může být složitý a měnit se v různých částech prostoru vlastností. Takovéto modely jsou flexibilní, což jim umožňuje zachytit složité vzorce v datech, avšak zároveň jsou náchylné na přetrénování (overfitting), zejména pokud jsou stromy příliš hluboké. Z tohoto důvodu jsou pro zlepšení generalizace a prevenci přetrénování používány různé techniky, předními z nichž jsou [9]:

- **Ořezávání (pruning)**
Rozhodovací strom je po vytvoření zredukován odebráním některých větví, čímž se zajistí menší variance alepší tak generalizační schopnosti.
- **Bagging**
Bootstrap Aggregating, zkráceně Bagging, spočívá ve vytváření několika rozhodovacích stromů na různých náhodně vybraných podmnožinách trénovacích dat a následném zprůměrování jejich výstupů. Tím se snižuje variance modelu a zvyšuje jeho robustnost vůči přetrénování.
- **Random Forest**
Podobně jako Bagging, také Random Forest kombinuje více rozhodovacích stromů vytvořených na různých náhodně vybraných podmnožinách trénovacích dat. Kromě toho také náhodně vybírá podmnožiny vstupních vlastností pro každý strom, což zajišťuje vyšší rozmanitost mezi stromy a tím zlepšuje generalizaci modelu.
- **Boosting**
Boosting je přístup, kdy jsou stromy trénovány sekvenčně a každý strom se zaměřuje na příklady, které byly předchozím stromem klasifikované jako chybné. To umožňuje kombinovat několik slabých modelů a tím vytvořit jeden silný.

3.1.8 Rozhodovací stromy s technikou gradient boosting

Gradient boosting je technika pro zlepšování přesnosti modelů rozhodovacích stromů prostřednictvím sekvenčního učení. Jde o rozšíření klasického boostingového přístupu, ve kterém jsou slabé modely přidávány jeden po druhém, přičemž každý další model je trénován

na chybách modelu předchozího. Směr největšího poklesu ztráty, neboli gradient, říká, jakým směrem je potřeba upravit výstup modelu, aby byla chybovost modelu co nejnižší. Jinými slovy, každý nový strom se snaží aproximovat chybu, které se dopustil součet všech předchozích stromů, a na jejím základě provést změny, které zaručí zlepšení prediktivních schopností modelu [9].

3.1.9 Užití rozhodovacích stromů

Praktické uplatnění rozhodovacích stromů doplňuje užití logistické regrese a nacházejí využití v mnoha oblastech díky své schopnosti modelovat složité interakce mezi proměnnými a díky jejich interpretovatelnosti. Mezi hlavní oblasti, kde se rozhodovací stromy používají, patří:

- **Doprava**

Uplatnění nachází rozhodovací stromy v dopravě například v predikci poptávky sdílených jízd. Studie prediktivních modelů předpovídajících poptávku ride-hailingové společnosti DiDi Chuxing, jež je předním provozovatelem těchto služeb v Číně, ukázala, že rozhodovací stromy s technikou gradient boosting dosáhly nejlepší přesnosti predikce [18].

- **Marketing**

Společnosti jako Netflix a Spotify používají rozhodovací stromy v rámci doporučovacích algoritmů. Na základě uživatelského chování, demografických informací a preferencí napomáhají tyto modely doporučovat relevantní obsah [21].

Díky své flexibilitě a spolehlivosti představují rozhodovací stromy silný nástroj v mnoha doménách. Jejich intuitivní struktura navíc umožňuje výsledky snadno interpretovat, a to i bez pokročilé znalosti strojového učení.

3.2 Shrnutí

Prediktivní modely jsou zásadním nástrojem pro analýzu historických dat a predikci budoucích výsledků a událostí s aplikací v mnoha oblastech. Korektní kategorizace potřebného výstupu umožňuje využití správného přístupu pro vyhodnocení historických informací, ať jde o kategorizaci dat, nebo predikci spojitých hodnot.

Statistické učení, nehledě na přítomnost supervize, je zaměřeno na rozpoznávání vzorců a vztahů mezi daty, což je nepostradatelné při tvorbě predikcí na základě informací z minulosti. Lineární a logistická regrese jsou hojně využívány k analýze vztahů mezi nezávislými a závislými proměnnými v mnoha oblastech, jako jsou finanční služby, zdravotnictví či doprava, nejen automobilová, nýbrž také cyklistická, kde umožňují například předpovídat vytíženost dopravních uzlů. Primárním cílem regresních modelů je predikovat na základě lineárních vztahů mezi proměnnými.

V případech, kdy jsou vztahy vstupních proměnných nelineární, jsou využívány rozhodovací stromy. Jejich aplikace umožňuje modelovat složitější vzorce ve vztazích vstupních vlastností, které by bylo obtížné zachytit tradičními regresními modely, načež zachovávají dobrou interpretovatelnost.

Kapitola 4

Analýza

Pro návrh spolehlivého a praktického řešení je nezbytné provést důkladnou analýzu uživatelských a technických požadavků a zohlednit jejich závažnost. Kromě toho je třeba prozkoumat služby, které nabízejí stávající řešení, a identifikovat oblasti pro jejich možná vylepšení.

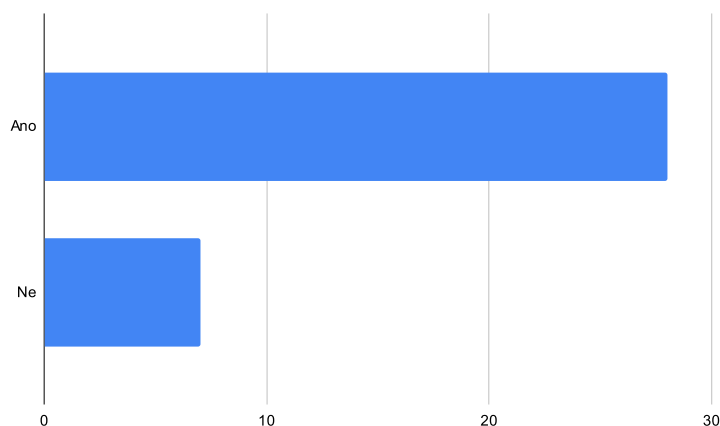
Hlavní uživatelskou skupinou, k jejímž potřebám by měl být nástroj přizpůsoben, je široká veřejnost. Tato skupina jako celek a její individuální subjekty mají specifické požadavky, které je třeba zohlednit při návrhu aplikace, aby poskytovala praktické a hodnotné řešení.

4.1 Uživatelská analýza

Průzkum mezi širokou veřejností byl realizován prostřednictvím dotazníku zaměřeného na pochopení preferencí a potřeb uživatelů, které jsou klíčové pro návrh řešení odpovídajícího jejich požadavkům. Dotazník byl distribuován mezi potenciální uživatele služeb sdílené cyklistiky a jeho cílem bylo získat náhled do následujících oblastí:

- **Zkušenosti uživatelů s nedostupností prostředků**

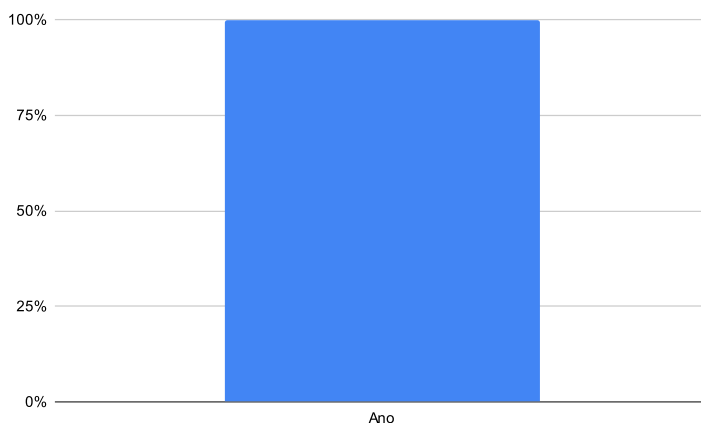
Uživatelé byli výběrem z možností dotázáni, zda-li se v minulosti setkali se situací, kdy v požadovaný čas nebyl na dané stanici přítomný žádný sdílený prostředek. Poměr odpovědí na dotaz je možné vidět na obrázku 4.1.



Obrázek 4.1: Zkušenost uživatelů s nedostupností prostředku

- **Názor na užitečnost funkce předpovědi dostupnosti prostředků v konkrétní datum a čas**

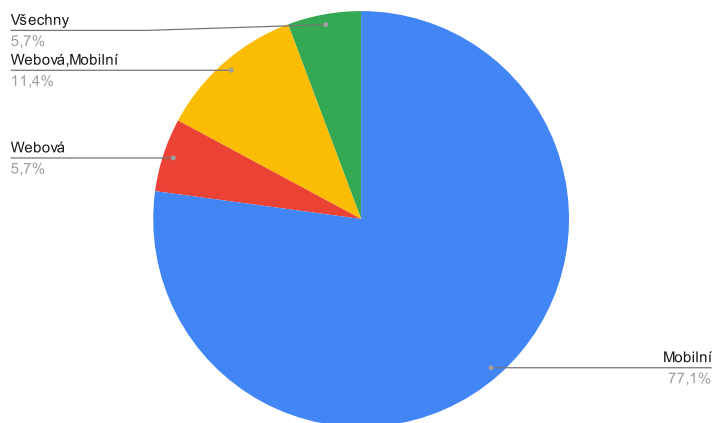
Dvojici možností nabízela odpověď na otázku "Zastáváte názor, že možnost zjistit dostupnost v konkrétní datum a čas v budoucnosti je užitečná funkce?", s přehledem vyjádření zkoumané skupiny na grafu 4.2.



Obrázek 4.2: Názor na užitečnost funkce predikce dostupnosti

- **Preferovaná forma aplikace**

Důležitým aspektem pro využitelnost aplikace je cílová platforma, pro kterou má být aplikace navržena a vyvinuta. Mezi možnostmi výběru jsou aplikace mobilní, desktopová, nebo webová. Kombinace odpovědí s jasnou převahou mobilní aplikace je znázorněna na grafu 4.3.



Obrázek 4.3: Preferovaná forma aplikace

Z celkového počtu 58 respondentů by celá tato skupina uvítala možnost zjistit dostupnost sdíleného kola dopředu. Rovněž je z takto provedené analýzy patrné, že většina uživatelů se přiklání k prezentaci výsledků prediktivní analýzy prostřednictvím mobilní aplikace.

4.2 Analýza stávajících řešení

Analytických nástrojů poskytujících statistická data služeb sdílené cyklistiky je celá řada, jednotlivá řešení přinášejí různé přístupy a pojetí těchto informací. Kromě nástrojů poskytovatelů, které prezentují data o aktuální dostupnosti prostředků sdílené cyklistiky na stanicích, existují také aplikace třetích stran popsané v kapitole 2.3. Mezi přední nedostatky řešení popsanych v této kapitole patří absence prognózy dostupnosti prostředku a nekompletní množina poskytovatelů služeb.

4.3 Funkční požadavky

Obecné požadavky vyplývají z principů a nepsaných pravidel uživatelských aplikací, přičemž důraz je kladen na intuitivní ovládání, přehledné uživatelské rozhraní a snadnou dostupnost klíčových funkcí bez zbytečné složitosti či nadbytečných kroků. Aplikace by měla být responzivní, optimalizovaná pro různé typy zařízení a poskytovat uživateli srozumitelnou zpětnou vazbu při interakci. Speciální požadavky na funkcionalitu konečného řešení odrážejí výsledky uživatelského průzkumu a analýzy aktuálně dostupných aplikací.

- **Zobrazení interaktivní mapy a vykreslení bodových prvků**

Bodové elementy reprezentující stanice zanesené do interaktivní mapy území města Brno je vhodný způsob reprezentace informací o aktuální i budoucí dostupnosti prostředků.

- **Klasifikace záznamů a výpočet pravděpodobnosti dostupnosti prostředku a následná prezentace výsledku**

Analýza historických dat pro predikci pravděpodobnosti dostupnosti prostředků v daném čase na konkrétních stanicích a vizuální prezentace výsledků ve srozumitelné podobě.

- **Dynamické aktualizace datových sad pomocí dostupných API**

Automatické stahování a synchronizace aktuálních dat o stanicích a dostupných prostředcích pomocí rozhraní API, čímž je zajištěna přesnost a aktuálnost informací.

Kapitola 5

Návrh

Navrhovaná aplikace se bude zaměřovat na analýzu dat a predikci dostupnosti prostředků služeb sdílené cyklistiky ve městě Brno. Cílem aplikace je poskytnout uživatelům interaktivní nástroje pro zjištění dostupnosti v plánované době jízdy. Následující kapitola je zaměřena na námět konkrétního řešení splňujícího požadavky popsané v předchozí kapitole. V příštích sekcích jsou představeny jednotlivé komponenty systému, jejich vzájemné vztahy a technologické prostředky použité k implementaci.

5.1 Architektura

Celé řešení bude sestávat ze dvou hlavních komponent - serverové aplikace shromažďující a transformující data poskytovatelů, a mobilní aplikace, určené pro vizualizaci dat získaných ze serveru pomocí komunikačního rozhraní.

5.1.1 Serverová aplikace a API rozhraní

Hlavní úlohou aplikace na straně serveru bude shromažďování a transformace dat poskytovatelů služeb sdílené cyklistiky. Tato data, zredukovaná na relevantní informace, budou ukládána do databáze periodicky nebo nárazově po transformaci v závislosti na datech konkrétních provozovatelů. Vzhledem k možnosti zjistit aktuální informace společnosti Nextbike prostřednictvím jejich API rozhraní, budou tato data obstarávána tímto způsobem ve vhodně zvolených intervalech. Absence API společnosti Rekola vyžaduje, aby byla data serverem podle potřeby vhodným způsobem transformována do stavu potřebného pro uložení do databáze.

Komunikace mezi serverovou a mobilní aplikací bude uskutečněna prostřednictvím API rozhraní na straně serveru. API bude obsahovat následující endpointy:

- **/stations**
Vrací seznam stanic obou společností. Každá položka seznamu obsahuje identifikátor, polohu danou zeměpisnou šířkou a délkou a poskytovatele, kterému stanice přísluší.
- **/stations/*id***
Pomocí identifikátoru stanice bude možné získat detail dané stanice, který bude, navíc od struktury stanice v seznamu stanic, obsah také název a podrobné informace o dostupnosti kol v podobě pravděpodobnosti nebo aktuální dostupnosti.

- **/stations/company/název společnosti**

Na základě názvu poskytovatele bude možné získat pouze seznam stanic provozovaných daným poskytovatelem.

- **/configuration**

Tento endpoint bude sloužit ke konfiguraci serveru prostřednictvím API rozhraní, především k načítání, transformaci a ukládání informací z datových sad poskytovatelů ze souborů. Dále bude touto cestou možné znovu trénovat prediktivní model.

Vedle základních skupin bude API rozhraní nabízet možnost zjistit informace o konkrétní stanici, jako je její název a podrobná data o dostupnosti kol, ať už aktuální nebo pravděpodobná. Dále bude serverová aplikace umožňovat získat seznam stanic pouze pro určitého poskytovatele.

5.1.2 Prediktivní model

Návrh modelu pro prediktivní analýzu je závislý na řadě faktorů. Vedle povahy vstupních proměnných a výpočetní náročnosti je důležitým z kritérií také reálný význam uživateli zobrazených informací. Z tohoto důvodu je z možností vyřazen model lineární regrese, jehož výstupem je spojitá hodnota, nikoliv procentuální pravděpodobnost. Konečný výběr je zúžen na následující možnosti:

- **Logistická regrese**

Omezit se pouze na binární klasifikaci nebo procentuální pravděpodobnost výskytu prostředku umožňuje logistické regrese. Model tedy vrací hodnotu mezi 0 a 1, určující míru jistoty přítomnosti prostředku na místě. Tato pravděpodobnost pak může být dále interpretována jako klasifikaci, zda-li bude kolo na stanici či, například pomocí zvoleného prahového parametru (např. 0,5). Pokud je výsledná hodnota vyšší než tento práh, model předpoví, že prostředek bude dostupný. Logistická regrese tak nabízí přímější a intuitivnější způsob prezentace výsledků predikce, který je lépe uchopitelný pro účel aplikace.

- **Multinomiální logistická regrese**

V případech, kdy je třeba rozlišovat mezi více než dvěmi kategoriemi, lze využít rozšířenou variantu – multinomiální logistickou regresi. Tento model je schopen určit pravděpodobnost výskytu více tříd současně, například nulového, malého, středního nebo vysokého počtu prostředků. Namísto jediné binární odpovědi tak poskytuje jemnější rozlišení situací a může být využit tam, kde je užitečné rozdělit výstup do více interpretačně odlišných úrovní dostupnosti.

- **Rozhodovací strom**

Za předpokladu, že je povaha vztahu vstupních vlastností nelineární, je žádoucí použít rozhodovací stromy. Jejich silnou stránkou je flexibilita, interpretovatelnost a díky technikám jako je ořezávání a boosting je docíleno velice dobré preciznosti těchto modelů.

Rozhodovací strom je vhodnější volbou oproti regresnímu modelu zejména kvůli povaze vstupních proměnných a vztahů mezi nimi. Proměnné jako roční období, den v týdnu či hodina v rámci dne vykazují nelineární a často i kategoriální charakter. Tyto proměnné nemusí mít lineární vliv na výstup a jejich interakce jsou často závislé na kontextu - například vliv hodiny během dne na dostupnost kola se může výrazně lišit mezi pracovními

dny a víkendy. Rozhodovací strom se s těmito typy proměnných vyrovnává přirozeně, bez nutnosti transformace, a umožňuje modelu automaticky rozpoznávat podmíněné závislosti pomocí větvení. Oproti tomu regresní modely předpokládají lineární vztahy a obvykle vyžadují rozsáhlé předzpracování kategoriálních dat (např. one-hot encoding), což může vést ke zbytečnému navýšení počtu proměnných a narušení přirozených vztahů mezi kategoriemi. Další výhodou rozhodovacích stromů je možnost jejich posílení pomocí techniky gradient boosting, která kombinuje více slabších modelů za účelem vytvoření výkonnějšího prediktoru. Tento přístup pomáhá omezit přetrénování a zvyšuje schopnost modelu zachytit složité, nelineární vztahy mezi proměnnými. Navíc při použití stromu jako binárního klasifikátoru lze přirozeně interpretovat výstupní hodnoty jako pravděpodobnost výskytu sledovaného jevu (např. dostupnosti kola), jelikož výstup leží v intervalu 0–1. Kombinace těchto vlastností činí stromové modely robustní a přesné při predikci dostupnosti ve sdíleném systému jízdních kol.

5.1.3 Mobilní aplikace

Úlohou klientské části výsledného řešení v podobě mobilní aplikace bude nejen prezentace výsledků predikční analýzy, ale i interaktivní práce s daty v reálném čase. Aplikace poskytne uživateli intuitivní přehled o aktuální a budoucí dostupnosti dopravního prostředku a umožní mu tak efektivní plánování cest.

Základními prvky mobilní aplikace budou:

- **Interaktivní mapa**

Základní prvek klientské aplikace, který vytvoří kontext pro stanice značené body a umožní výběrem dané stanice zjistit aktuální nebo pravděpodobnou dostupnost.

- **Vizualizace predikce**

Dialogové okno, dostupné po zvolení konkrétní stanice, umožní uživateli po zadání data a času zjistit dostupnost prostředku v požadovanou dobu dopředu.

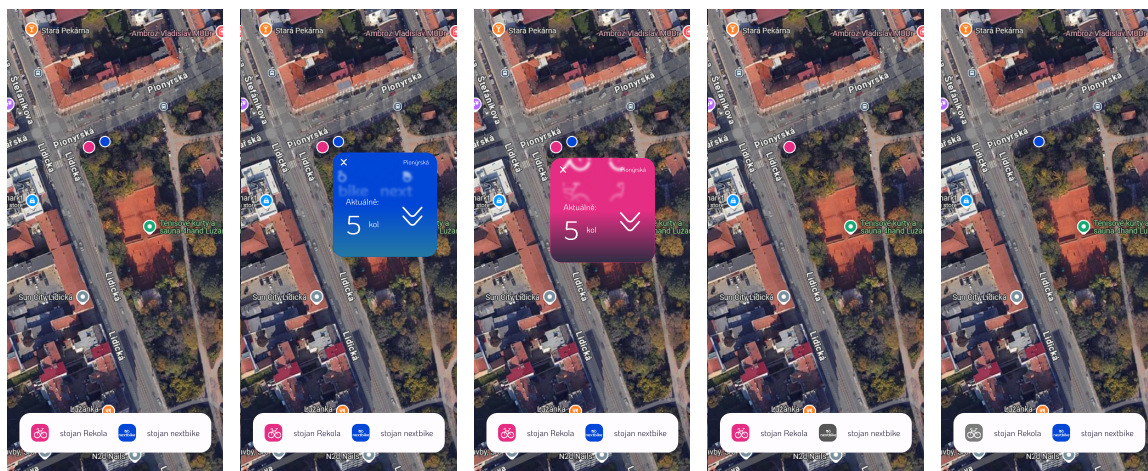
Obě tyto části budou realizovány prostřednictvím technologie .NET pro snadnou škálovatelnost a podporu a také pro možnost tvorby multiplatformní klientské části řešení díky .NET MAUI, zefektivňující tak vývoj a reparabilitu. Tato technologie disponuje také knihovnou ML.NET, díky které je velice snadné implementovat prediktivní modely, jako jsou logistická regrese nebo rozhodovací stromy. Výsledkem tedy bude technologií sjednocené a snadno spravovatelné řešení.

5.2 Vstupní data

Hlavními zdroji vstupních informací jsou API společnosti Nextbike a surové datové sady společnosti Rekola, které zahrnují informace o jízdách identifikovaných počátečními a cílovými body a časovými značkami začátku a konce cesty, které jsou pro výsledné řešení stěžejní. Konkrétní formát a struktura API rozhraní jsou popsány v kapitole 2.2.

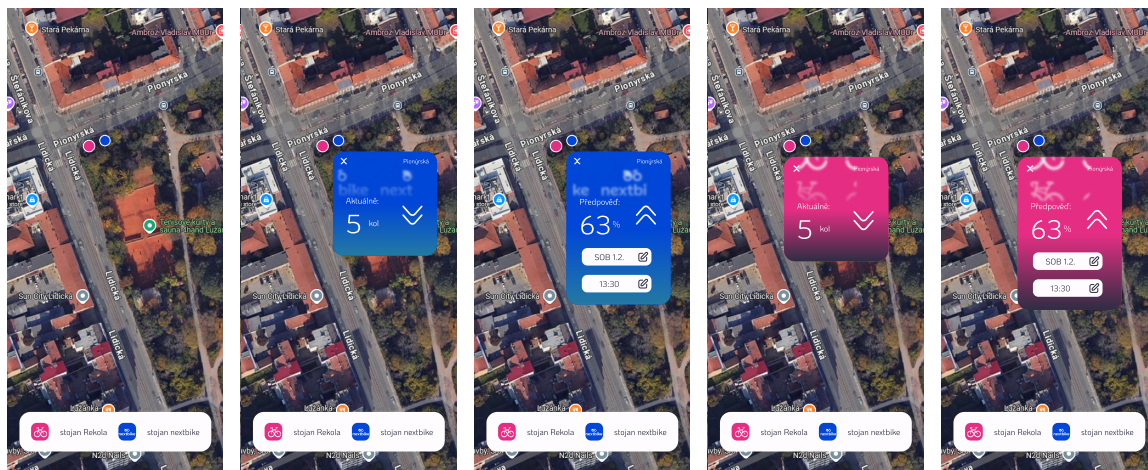
5.3 Uživatelské rozhraní

Hlavní obrazovka klientské aplikace bude sestávat z interaktivní mapy, do které budou zaneseny jednotlivé stanice reprezentované body s barevnou indikací v závislosti na provozovateli stanice. Detail stanice bude, v případě provozovatele Nextbike, obsahovat informace o aktuálním počtu dostupných prostředků na dané stanici, v případě společnosti Rekola údaj o pravděpodobnosti výskytu prostředku v daný moment. K dispozici také bude uživateli filtr pro vykreslení stanic jednoho nebo obou provozovatelů 5.1.



Obrázek 5.1: Interaktivní mapa s grafickými prvky

Prezentovat předpověď přítomnosti prostředku v datu a čase vybraném uživatelem bude mít za úkol rozšíření detailu stanice. Návrh modulu pro zjištění pravděpodobnosti dostupnosti prostředku je vyobrazen na obrázku 5.2.

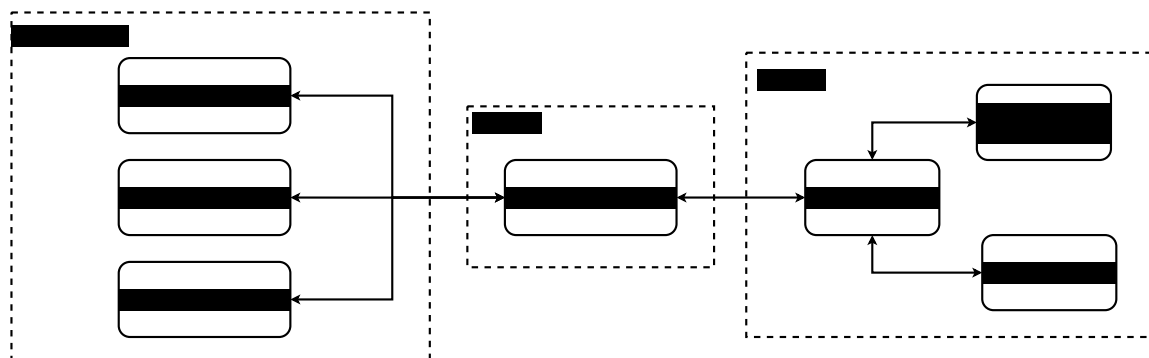


Obrázek 5.2: Interaktivní mapa s vizualizací zjištění pravděpodobnosti

Kapitola 6

Implementace

V následující kapitole je podrobně popsána implementace aplikace, jejíž návrh je popsán v kapitole 5. V sekcích této kapitoly jsou popsány realizace jednotlivých částí - backendové a frontendové - jejichž architektura je názorně zobrazena na obrázku 6.1. Dále je zde popsán vývoj prediktivního modelu a proces výběru vhodné kombinace modelů.



Obrázek 6.1: Architektura aplikace

6.1 Backend

Backendová část systému zajišťuje komplexní správu dat, od jejich získávání, přes transformaci, až po zpřístupnění klientské části aplikace. Dále je zodpovědná za provádění prediktivní analýzy dostupnosti sdílených kol.

Data jsou pravidelně získávána z veřejných API poskytovatelů sdílené cyklistiky a ukládána do databáze ve strukturované podobě. Součástí backendu je i logika pro transformaci a zpracování těchto dat, která je převádí a připravuje pro další využití, například pro vizualizaci na mapě či pro predikci. Struktura těchto dat a jejich vzájemné vztahy jsou zachyceny v databázovém schématu v podobě ERD, které je vyobrazeno na obrázku 6.2.

`Stations`	
PK	CONSTRAINT `PK Stations` (`Id`)
`Id` char(36) COLLATE ascii_general_ci NOT NULL	
`Uid` int NOT NULL	
`Name` varchar(255) CHARACTER SET utf8mb4 NO	
`Latitude` double NOT NULL	
`Longitude` double NOT NULL	
`Company` int NOT NULL	

`Availabilities`	
PK	CONSTRAINT `FK Availabilities Stations Station
PK	CONSTRAINT `PK Availabilities` (`Id`)
`Id` char(36) COLLATE ascii_general_ci NOT NULL	
`DateAndTime` datetime(6) NOT NULL	
`BikeCount` int NOT NULL	
`StationId` char(36) COLLATE ascii_general_ci NOT	
`TemperatureAtDateAndTime` float NOT NULL	
`RainProbabilityAtDateAndTime` float NOT NULL	

Obrázek 6.2: Databázové schéma aplikace

Pro předpověď dostupnosti kol na konkrétní stanici a v konkrétním čase je využit model rozhodovacího stromu. Tento model vyhodnocuje historická data o využívání stanic a bere v potaz i kontextové informace, jako je den v týdnu nebo počasí. Na základě těchto vstupů je schopen s určitou mírou pravděpodobnosti určit, zda budou na stanici dostupná kola. Podrobný popis prediktivního modelu je uveden v části 6.2.

Backend tak tvoří analytické jádro celé aplikace, které poskytuje klientovi nejen aktuální data, ale i prediktivní vhled pro plánování využití sdílených kol.

6.1.1 API rozhraní

Přístup k funkcím backendové aplikace a informacím o stanicích je uskutečněn pomocí API rozhraní. Tímto způsobem jsou realizovány jak dotazy na seznamy a detaily stanic, tak konfigurační funkce aplikace. Backendová část aplikace poskytuje následující endpointy:

- **/stations**
Endpoint vrací seznam všech stanic.
- **/stations/by-company?company={company}**
Vrací seznam stanic podle vybraného provozovatele.
- **/stations/{stationId}**
Vrací detailní informace o stanici na základě jejího identifikátoru ('GUID').
- **/stations/{stationId}/availability?datetime={datetime}**
Vrací predikovanou dostupnost kol na konkrétní stanici k zadanému datu a času.
- **/stations/search?name={name}**
Vyhledává stanice podle názvu. Výsledkem je seznam stanic obsahujících hledaný text v názvu.
- **/configuration/Nextbike/load-stations**
Načte a uloží seznam stanic od provozovatele Nextbike.
- **/configuration/Nextbike/load-availabilities**
Načte a uloží historická data o dostupnosti kol od provozovatele Nextbike.

- **/configuration/rekola/load-stations**
Načte a uloží seznam stanic od provozovatele Rekola.
- **/configuration/rekola/load-availabilities**
Načte a uloží historická data o dostupnosti kol od provozovatele Rekola.
- **/configuration/train**
Spustí trénink prediktivního modelu pro všechny provozovatele.
- **/configuration/train/{company}**
Provede trénování prediktivního modelu pouze pro vybraného provozovatele.

API rozhraní tvoří prostředníka mezi backendovou a frontendovou částí aplikace. Obstarává požadavky nejen na prediktivní analýzu od uživatele a poskytuje klientské aplikaci výsledky.

6.1.2 Open Meteo API

K získání dat o teplotních průměrech a průměrech procentuální pravděpodobnosti srážek v rámci dnů bylo použito Open Meteo API ¹, jelikož podporuje bezplatné získání informací o počasí až od roku 1980 v rámci jednotlivých dnů.

6.2 Prediktivní model

Během implementace jsem pracoval s několika variantami kombinací prediktivních modelů pro jednotlivé poskytovatele. Výsledná implementace je podrobně popsána v sekci 6.2.1. V průběhu vývoje jsem pracoval s následujícími verzemi:

1. Společný prediktivní model pro oba poskytovatele

První variantou implementace prediktivního modelu byl jeden společný model pro oba poskytovatele. Tato varianta však poskytovala zavádějící výsledky, vzhledem k tomu, že datová sada společnosti Rekola neobsahuje vyvažující hodnoty.

2. Individuální prediktivní modely pro každého z poskytovatelů

Další implementovanou variantou byly dva rozdílné modely pro jednotlivé poskytovatele. Toto umožnilo sledovat vliv dat společnosti Rekola, která jsou nevyvážená a poskytují pouze pozitivní záznamy, na celkovou spolehlivost a přesnost celého modelu. Tato cesta umožnila model pro stanice společnosti Nextbike spolehlivě předpovídat pravděpodobnost, což ale nebyl případ pro stanice společnosti Rekola.

Pro účely výsledné aplikace byla vybrána kombinace dvou rozhodovacích stromů separátních pro každého poskytovatele. Tato varianta umožnila spolehlivě předpovídat dostupnost na stanicích Nextbike a poukázat na vliv rozdílných datových sad mezi poskytovateli.

¹<https://open-meteo.com/>

6.2.1 Implementace a trénování

Prediktivní modely jsou implementovány a natrénovány pomocí knihovny ML.NET ² na historických datech dostupnosti na stanicích. Výsledkem jsou binární klasifikátory, které zjišťují pravděpodobnost dostupnosti prostředku na stanici daného poskytovatele.

Rozhodovací stromy využívají algoritmus FastTree implementující strom s technikou gradient boosting. Tyto modely, natrénované na záznamech o počtu prostředků na stanicích, odhadují pravděpodobnost, že bude na daném místě v požadovaný čas dostupný alespoň jeden dopravní prostředek, a to na základě následujících prediktorů:

- **StationId**
Jednoznačný identifikátor stanice.
- **Temperature**
Teplotní průměr v rámci dne nebo teplota v hodinu, kdy byl záznam o dostupnosti pořízen.
- **RainProbability**
Průměr pravděpodobnosti deště v rámci dne nebo pravděpodobnost srážek v hodinu, kdy byl záznam pořízen.
- **Weekday**
Den v týdnu, v rámci kterého byl záznam pořízen.
- **Season**
Roční období, v rámci kterého byl záznam pořízen.
- **HourOfDay**
Hodina, ve kterou byl záznam pořízen.
- **BikeAvailability**
Binární klasifikace dostupnosti prostředku na stanici. Konkrétní počet kol záznamu je binárně klasifikován a říká, zda byl alespoň jeden prostředek na stanici dostupný, či nikoliv.

Takto zvolené prediktory umožní modelu sledovat sezónní, týdenní a denní trendy a vliv počasí na dostupnost dopravních prostředků. Výsledné modely tak poskytují praktické predikce dostupnosti prostředků a napomáhají tak spolehlivému plánování.

²<https://dotnet.microsoft.com/en-us/apps/ai/ml-dotnet>

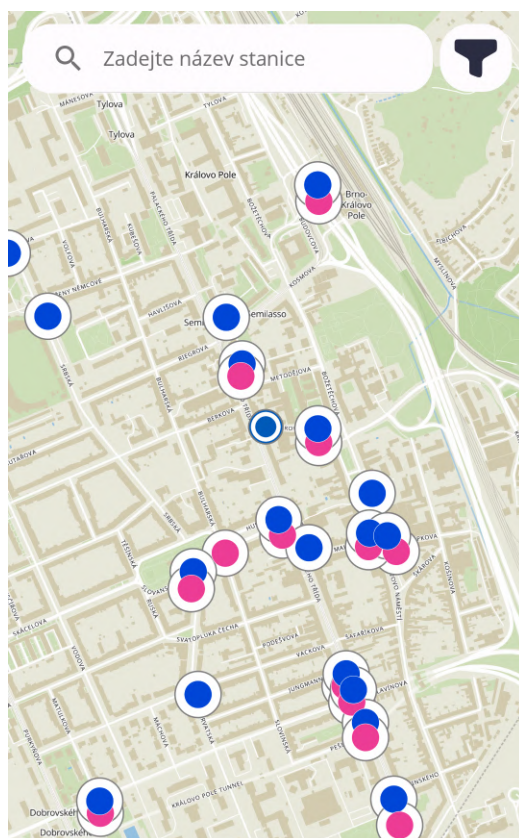
6.3 Frontend

Frontendová část systému slouží jako uživatelské rozhraní pro zobrazování dat z backendové části a interakci uživatele s aplikací. Pro jeho implementaci byl zvolen framework .NET MAUI (Multi-platform App UI), který umožňuje vývoj multiplatformních aplikací s jednotným kódem pro různé cílové platformy (Android, iOS, Windows, macOS). Tento přístup umožňuje efektivní vývoj a snadnou údržbu aplikace.

Uživatelské rozhraní aplikace je navrženo s důrazem na přehlednost a intuitivní ovládání. Sestává z několika klíčových komponent, které umožňují uživateli pohodlný přístup k informacím o sdílených kolech a stanicích a jejich analýzám. Mezi hlavní komponenty patří:

- **Mapa a seznam stanic**

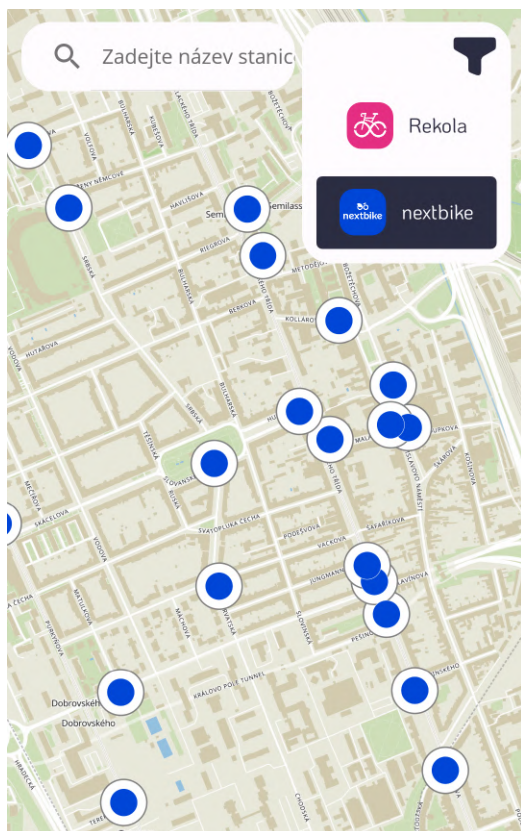
Interaktivní mapa tvoří hlavní prostředí aplikace a základní vizuální kontext pro zobrazení stanic a jejich polohy. Pro implementaci byla zvolena knihovna **Mapsui**, která je open-source, dobře rozšiřitelná a podporuje různé mapové zdroje. Mapa umožňuje zobrazovat jednotlivé stanice pomocí vlastních ikon, reaguje na doteky uživatele (např. kliknutí na stanici) a je úzce propojena s ostatními komponentami – výběr stanice na mapě zobrazí její detail, aplikovaný filtr dynamicky mění zobrazený obsah a vyhledávací pole umožní rychle vyhledat konkrétní stanici. Tím je dosaženo intuitivního a interaktivního uživatelského zážitku. Podoba mapy a do ní zasazených bodů je vyobrazena na obrázku 6.3.



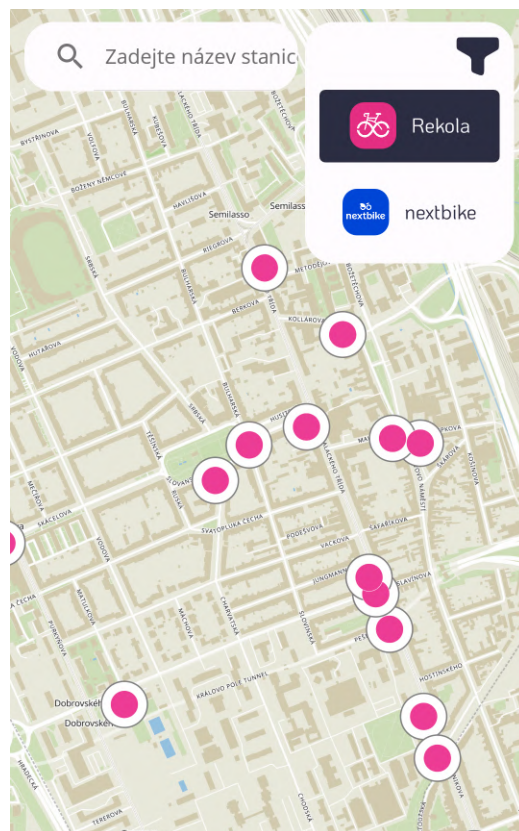
Obrázek 6.3: Interaktivní mapa s body reprezentující stanice

- **Filtr poskytovatele služeb**

Pomocí tlačítka pro filtrování su uživatelé mohou zobrazit nabídku poskytovatelů. Při výběru jedné z těchto stanic kliknutím se v mapě zobrazí pouze stanice provozované danou společností. Rovněž také filtrování ovlivňuje textové vyhledávání pomocí vyhledávacího pole, kdy se ve výsledcích zobrazí pouze stanice vybrané společnosti. Podobu mapy při aplikaci filtru pro společnost Nextbike, resp. Rekola, je možné vidět na obrázcích 6.4 a 6.5.



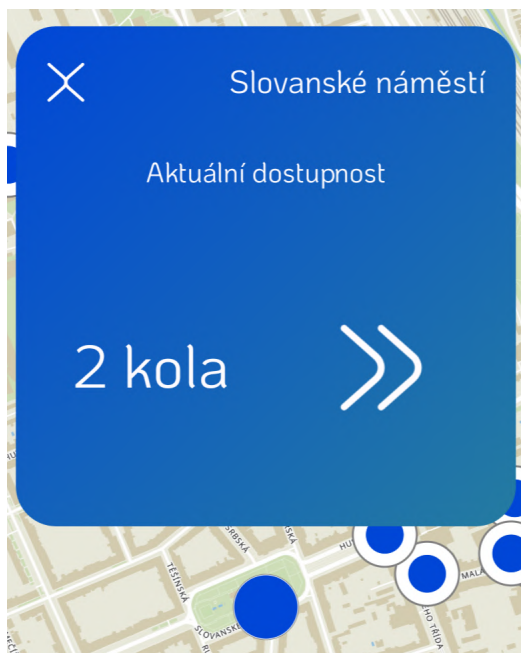
Obrázek 6.4: Filtr stanic poskytovatele Nextbike



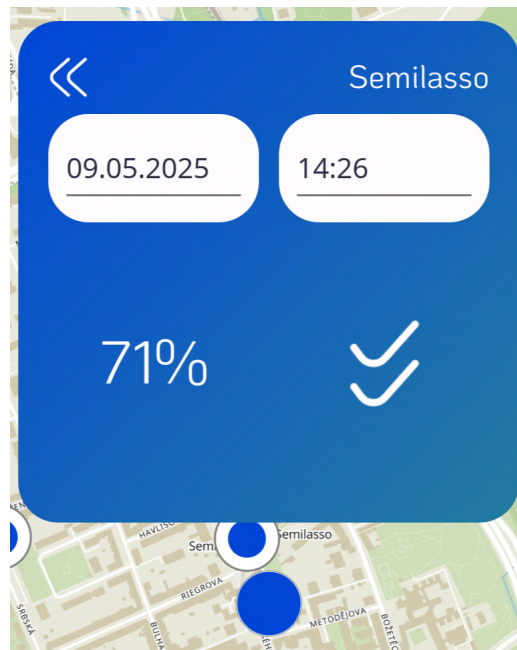
Obrázek 6.5: Filtr stanic poskytovatele rekola

- **Detail stanice**

Detail stanice v podobě vyskakovacího okna prezentuje podrobné informace o zvoleném místě. Na první ze dvou stran je zobrazen celý název stanice, aktuální počet kol, respektive předpovězená dostupnost prostředku v daný moment. Příslušnost k provozovateli je označena barevným pozadím celého prvku. Druhá strana umožňuje uživateli zadat datum a čas pomocí vstupních polí, pro které je požadováno zjistit pravděpodobnost dostupnosti. Tato informace, spolu s celým názvem stanice, druhou stranu okna doplňuje. Obě části okna, zobrazené na obrázcích 6.6 a 6.7, navíc obsahují ovládací prvky v podobě ikon, pomocí kterých může uživatel mezi pohledy navigovat, potvrdit výběr data a celou komponentu zavřít. Otevření detailu je při výběru stanice navíc doplněno o plynulý přechod mezi aktuálním zobrazením na vycentrování a přiblížení bodu.



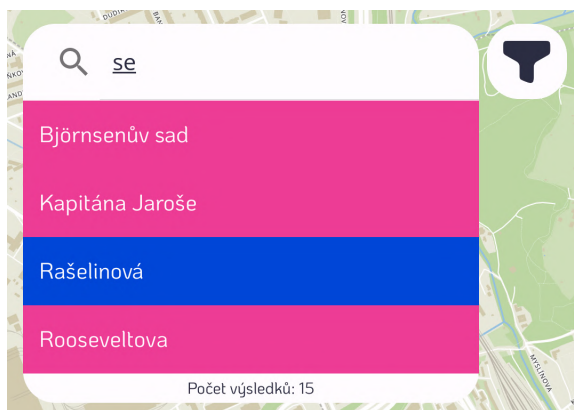
Obrázek 6.6: Detail stanice Nextbike



Obrázek 6.7: Predikce dostupnosti

- **Vyhledávací panel**

Možnost rychle najít stanici podle názvu umožňuje vyhledávací pole. Po zadání hledaného textu nebo jeho části se uživateli zobrazí 4 položky spolu s celkovým počtem stanic, jejichž název vyhledávaný řetězec obsahuje. Kliknutím na stanici z výběru je uživatel přenesen do detailu stanice. Skutečnou podobu komponenty je možné vidět na obrázku 6.8.



Obrázek 6.8: Výsledek vyhledávání názvu stanice

- **API klient**

Pro komunikaci s backendovou částí aplikace slouží vlastní API klient, který zajišťuje volání HTTP endpointů a zpracování odpovědí. API klient je navržen tak, aby byl snadno testovatelný, přehledný a rozšiřitelný o další datové zdroje. Podporuje získávání aktuálních i předpovídaných dat o dostupnosti kol, a umožňuje zadání parametrizovaných dotazů včetně filtrů poskytovatelů nebo výběru konkrétní stanice.

6.4 Problémy při implementaci

Během implementace navrhované aplikace jsem se setkal s řadou problémů, některé z nichž vývoj aplikace určitým směrem zneprůjemnil, jiné přímo znemožnily zrealizovat některé části funkcionality aplikace navrhované v kapitole 5. Níže jsou popsány hlavní překážky při vývoji řešení.

6.4.1 Nekonzistence datasetů

První z překážek při implementaci tvořila nekonzistence dat mezi poskytovateli. Společnosti provozující služby sdílené cyklistiky přistupují rozdílným způsobem k zpřístupnění informací - zatímco společnost Nextbike umožňuje data získat výhradně v reálném čase pomocí API rozhraní, kdy je možné získat stavy jednotlivých stanic z pohledu obsazenosti, provozovatel Rekola poskytuje informace po domluvě z oficiálního archivu v podobě záznamů o jízdách. Tyto odlišnosti si vyžadují speciální přístup k získání a transformaci získaných dat.

Další z nepříjemností je nekonzistence dat v rámci poskytovatele Rekola. Poskytnuté datasety obsahují informace v různých strukturách a různých formátech datových typů, konkrétně časových značek jízd. Důsledkem toho vznikla potřeba specifickým způsobem předchystat data pro následnou transformaci a uložení do databáze.

Informace o jízdách navíc netvoří vhodnou reprezentaci dostupnosti na stanicích v časových úsecích. Po transformaci je databáze ochuzena o vyvažující hodnoty pro dostupnost, a to jsou data v časech, kdy na stanici nebylo kolo dostupné. V hodinových intervalech, ve kterých se pro danou stanici nenachází záznam o počtu prostředků, tak bylo nutné vytvořit tento záznam. Tímto domyšlivým způsobem ale došlo k situaci, že takto doplněná data nejsou vhodně reprezentativní z důvodu poměrně nízké četnosti výpůjček.

6.4.2 Mapová knihovna

Za účelem eliminace finančních nákladů na vývoj a provoz aplikace vznikla potřeba obcházet některé přímočařejší možnosti na náročnost nákladnějšími alternativami. Pro implementaci interaktivní mapy v klientské části aplikace lze použít balíček *.NET MAUI Maps*, s jehož použitím je implementovat mapu poměrně jednoduché, protože ale využívá map nativních pro danou platformu - Google Maps na zařízeních Android, resp. Apple Maps na operačních systémech iOS - nebylo tak možné bezplatně využít tuto knihovnu, neboť společnost Google zavádí jisté omezení pro počet dotazů na zobrazení a navigaci na mapě. Tyto podmínky nebyly omezující v rámci implementace aplikace, ale pro následný provoz tvoří riziko pro provozuschopnost.

Za účelem řešení tohoto problému byla pro implementaci vybrána knihovna *Mapsui*, která sice nevyžaduje poplatky na provoz, má ale taky svá omezení. Jedním z nich je chybějící podpora vektorových map, uživateli je možné tedy zobrazit mapu pouze v rastrové podobě, což v určitých případech může snížit její čitelnost. Další závažnější překážkou je jednoduchost implementace vizuálních vlastností bodových prvků. Knihovna obsahuje pouze jeden styl pro implementaci bodu značícího stanici, tudíž bylo potřeba, pro co nejprůběžnější uživatelský zážitek a dodržení grafického návrhu aplikace, body stanic vykreslovat vlastním způsobem.

Kapitola 7

Testování

Následující kapitola se zabývá testováním předpovědi dostupnosti za použití prediktivního modelu a klientské části aplikace na skupině uživatelů. Testování probíhalo již během vývoje běžným používáním a laděním, a podrobné přístupy k této problematice budou popsány v následujících sekcích.

7.1 Testování prediktivních modelů

Testování prediktivního modelu pro poskytovatele Rekola nebylo z důvodu podoby datasetu vhodné. Testován byl tedy pouze model pro společnost Nextbike, a to formou vytvoření testovací sady obsahující předpovědi dostupnosti prostředků na různých stanicích ve vlnách identifikovaných časovou značkou. Tyto předpovědi pak byly porovnány s reálnými stavy stanic.

Testovací sada se skládala z předpovědí pro šestici vybraných stanic v různých dnech a časech. Následuje ukázka predikcí a reálných stavů stanic, které byly předpovězeny a získány dne 2.5.2025 v 15:00:

- **Stanice Moravské náměstí - Čechova**
Predikovaná dostupnost alespoň jednoho prostředku byla 62%, ve skutečnosti se jich na stanici nacházelo 8.
- **Stanice Moravské náměstí - Jostova**
Ve zmiňovanou dobu se na stanici nacházel 2 prostředek. Předpovídaná dostupnost byla 54%.
- **Stanice Jakubské náměstí**
S pravděpodobností 52% model předpověděl dostupnost alespoň jednoho prostředku na stanici, skutečný stav byl 5 kol.
- **Stanice Sirotkova - Parkovací stání**
Na této stanici model předpověděl výskyt prostředku na 43%, na stanici se v danou dobu žádné kolo nenacházelo.
- **Stanice Oblá - UMC Nový Lískovec**
Výsledkem předpovědi na této stanici byla pravděpodobnost 24% výskytu sdíleného jízdního kola. Stanice byla v tuto dobu prázdná.

- **Stanice Semilasso**

Na stanici Semilasso byla předpovezena dostupnost pouze 34%, přesto se na stanici nacházely 4 prostředky.

Maximální nejistota výskytu prostředku byla stanovena hranicí 50%, pravděpodobnost vyšší než tato prahová hodnota tedy vyjadřuje jistotu přítomnosti jízdního kola v relativní míře k rozdílu hodnoty od prahu. Hodnota nižší než tento práh pak vyjadřuje situaci opačnou se stejnou relativní jistotou.

Dvojice predikcí a reálných stavů pak byly rozděleny do následujících skupin:

- **SP (Skutečně pozitivní)**

Případ, kdy model správně predikoval přítomnost kola a skutečně bylo alespoň jedno kolo k dispozici.

- **SN (Skutečně negativní)**

Model správně predikoval absenci kola a žádné kolo na stanici nebylo.

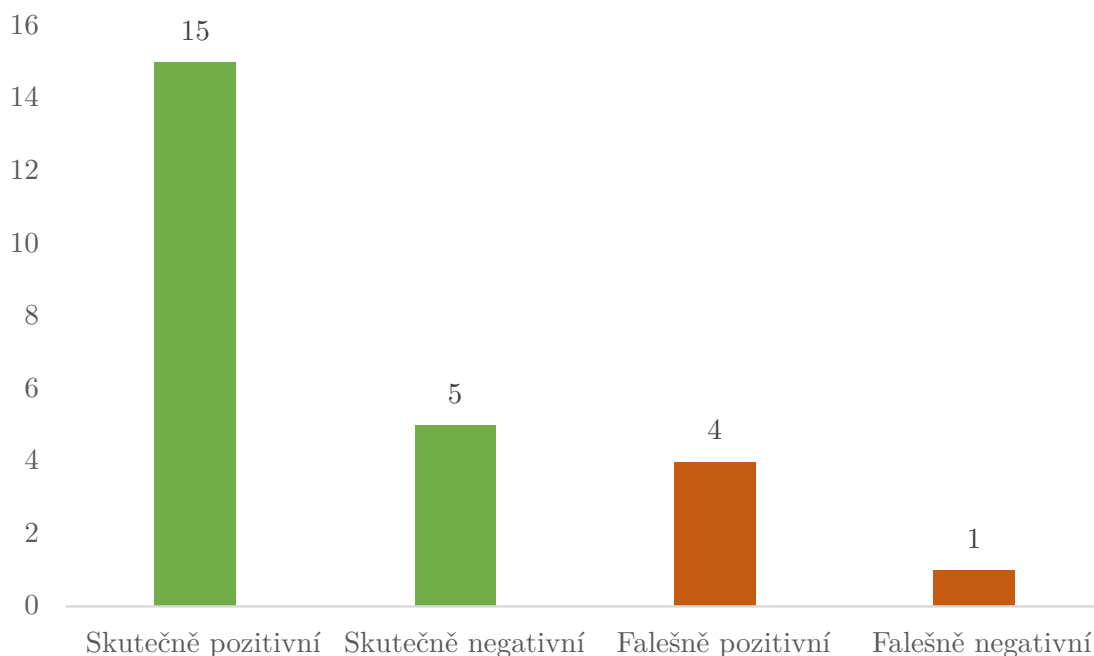
- **FP (Falešně pozitivní)**

Model predikoval přítomnost kola, ale žádné kolo se ve skutečnosti na stanici nenacházelo.

- **FN (Falešně negativní)**

Model nepředpokládal dostupné kolo, ale ve skutečnosti bylo alespoň jedno k dispozici.

Podrobné výsledky testování prediktivního modelu na množině 25 prvků testovací sady jsou znázorněny na grafu 7.1.



Obrázek 7.1: Výsledky testování prediktivního modelu

Z takto vytvořené sady predikcí a odpovídajících reálných stavů byla následně provedena evaluace úspěšnosti modelu s použitím tradičních metod následovně:

- **Pozitivní predikční hodnota (Precision):**

$$\text{Precision} = \frac{SP}{SP + FP} = \frac{15}{15 + 4} = \frac{15}{19} \approx 0,789 \quad (78,9 \%)$$

- **Přesnost (Accuracy):**

$$\text{Accuracy} = \frac{SP + SN}{SP + SN + FP + FN} = \frac{15 + 5}{15 + 5 + 4 + 1} = \frac{20}{25} = 0,800 \quad (80,0 \%)$$

- **Citlivost (Recall):**

$$\text{Recall} = \frac{SP}{SP + FN} = \frac{15}{15 + 1} = \frac{15}{16} \approx 0,938 \quad (93,8 \%)$$

- **F1-skóre:**

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot 0,789 \cdot 0,938}{0,789 + 0,938} \approx 0,859 \quad (85,9 \%)$$

V oblasti sdílené cyklistiky, kde se na výsledky předpovědi prediktivního modelu neklade takový důraz jako například v medicíně, kde se takovéto modely mohou využívat k odhalení nemoci na základě symptomů, lze z relativně malé sady testovacích dat vyvodit závěr, že model s F1-skórem přes 85% je dostatečně přesný a v praxi použitelný.

7.2 Testování uživatelského rozhraní

Uživatelské testování probíhalo na skupině subjektů s cílem obdržet odezvu na intuitivitu aplikace a celkovou použitelnost a srozumitelnost konceptu celého řešení. Uživatel dostal pouze minimální instruktáž, kdy úkolem testovaných subjektů bylo projít funkcemi aplikace, navigovat v ní do všech zobrazovaných prvků. Nastínění základního konceptu před testem a následná odezva uživatelů byly provedeny textovou korespondencí s testovacími subjekty.

7.2.1 Uživatel 1

Prvním testovacím subjektem je 20letý student informačních technologií. Jako sportovec se s aplikacemi pro sdílenou cyklistiku a problémem, který tato aplikace řeší, setkává na týdenní bázi. Níže jsou popsány hlavní body odezvy:

- Otáčení s mapou není možné a při orientaci v městském prostředí je tak identifikace potřebné stanice mírně těžší.
- Význam procentuálního vyjádření dostupnosti u stanic poskytovatele Rekola a skalární hodnoty u poskytovatele Nextbike uživateli zpočátku nebylo jasné, v konečném důsledku ale funkcionalitu pochopil.
- Tlačítka pro otevření karty zjištění dostupnosti a potvrzení výběru data a času nemají popis a jejich význam byl při prvním použití trochu nejasný.
- Celé řešení je intuitivní a disponuje minimalistickým designem bez zbytečných komponent.

7.2.2 Uživatel 2

Druhým uživatelem zapojeným do testování je 22letý student medicíny. Se službami sdílené cyklistiky se setkává velmi sporadicky, tudíž je jeho odezva k vizuálním vlastnostem aplikace váženější než odezva k funkčním požadavkům.

- Stejně jako testovací subjekt č.1 byl také druhý uživatel zpočátku zaskočen rozdílným způsobem prezentace dostupnosti v závislosti na poskytovateli, po krátkém seznámení se s aplikací se ale s významem srozuměl.
- Zvolený styl považuje za minimalistický, moderní a celkový koncept dobře uchopitelný.
- V některých případech se lze přes detail stanice prokliknout do jiného.
- Navzdory pevným hodnotám přiblížení a pozice při otevření detailu stanice relativně k jejímu bodu pomáhají přechody neztratit orientaci na mapě.

7.3 Závěr testování

Testování prediktivních modelů bylo smysluplné pouze pro model předpovídající dostupnost pro stanice společnosti Nextbike, jelikož poskytovatel služeb Rekola neposkytuje datasety užitečné pro prediktivní analýzu. Sadou testovacích vstupů byly zjištěny metriky spolehlivosti, preciznosti, přesnosti a F1-skóre modelu. Z důvodu poměrně omezené rozmanitosti funkcí mobilní aplikace se všichni uživatelé setkávali s podobnými problémy. V konečném důsledku je však koncept celého řešení, jak vyplynulo z korespondence s testovanými subjekty, dobře pochopitelný a zpracování aplikace přehledné a uživatelsky přívětivé.

7.4 Budoucí vývoj

Aplikace v současné podobě slouží jako nástroj pro zjištění aktuální nebo předpokládané dostupnosti prostředků výhradně na území města Brna. Jak serverová, tak klientská část aplikace však nabízí prostor pro další rozvoj a rozšíření funkcionalit.

- **Rozšíření do dalších měst**
Aplikaci lze rozšířit na další města, která podporují otevřená data o sdílených dopravních prostředcích. To by mohlo výrazně zvýšit její užitečnost a dosah.
- **Lepší integrace s poskytovateli**
Z důvodu nespolehlivé prognózy modelu pro společnost Rekola vzniká pro opravu tohoto stavu potřeba datové sady lépe vyhovující potřebám řešení. Možnost obsadit data pro poskytovatele Rekola stejným způsobem jako je tomu u společnosti Nextbike, tj. prostřednictvím API rozhraní, by výrazným způsobem přispěla ke kompletnosti aplikace.
- **Učení s pouze pozitivními vzorky**
Pro případ pouze pozitivních informací, které lze obdržet z datasetů společnosti Rekola, je možné pro výpočet pravděpodobnosti výskytu prostředku využít metodu Positive-Only učení, která zkoumá míru anomálie, která lze interpretovat jako nepřítomnost prostředku.

- **Styly map**

Díky využití knihovny Mapsui, která načítá mapové podklady z URL, je možné snadno implementovat podporu pro různé styly map, například satelitní nebo noční režim.

- **Více poskytovatelů**

Vedle sdílených kol by bylo možné rozšířit funkcionalitu i o jiné formy sdílené dopravy, například sdílené koloběžky. To by umožnilo komplexnější pohled na sdílenou dopravu ve městě.

- **Nápověda na použití klientské aplikace**

Z uživatelského testování vzešla potřeba lépe popsat funkce a důvod odlišného přístupu k různým poskytovatelům. Tento problém je možné vyřešit nápovědou například v podobě vyskakovacího okna nebo úvodní obrazovkou provázející uživatele funkcemi aplikace.

- **Kompletace datové sady společnosti Nextbike**

V době realizace tohoto řešení byla trénovací sada tvořena daty z necelých 5ti měsíců. Tímto je zaručeno ne příliš přesné předpovídání zejména například v zimních měsících, během kterých je pravděpodobné, že se dostupnost prostředků bude výrazně lišit. Za předpokladu, že by datová sada obsahovala informace v rámci alespoň jednoho kalendářního roku, by byla trénovací sada dostatečně obsáhlá a predikce tak dostatečně precizní.

Kapitola 8

Závěr

Sdílená mobilita, jakožto oblast stále rozvíjející se nejen v městech, reaguje na rostoucí požadavky na udržitelnou dopravu. Se stále zvyšujícím se využitím těchto služeb, mezi které patří mimo jiné také sdílená cyklistika, roste také potřeba efektivních nástrojů napomáhajících ke zkvalitnění těchto služeb. Dosud oblast sdílené cyklistiky neoplývá nástrojem nabízejícím prediktivní analýzu těchto služeb. V tomto kontextu vznikla tato práce, jež měla za cíl vytvořit uživatelsky přívětivý nástroj pro analýzu dat služeb pro sdílení kol ve městě Brno, který inovativním způsobem rozšiřuje možnosti současných systémů o prediktivní analýzu dostupnosti. Výsledkem je dvojice klient-server aplikací, prostřednictvím kterých se uživatelé dostává možnost zjistit dostupnost dopravního prostředku v budoucnosti. Této formy analýzy je dosaženo s použitím modelu rozhodovacího stromu. Samotné implementaci předcházelo podrobné seznámení se s aktuálním stavem sdílené cyklistiky v Brně a s nástroji poskytujícími různé formy analýzy, průzkum a ujasnění možností a limitů různých statistických modelů a následný návrh výsledné aplikace. Aplikace byla již při implementaci průběžně testována, následně pak byla podrobena nezávislému testování skupinou testovacích subjektů, kteří tvoří potenciální uživatelskou základnu pro výsledné řešení.

Implementovaná aplikace tvoří komplexní analytický nástroj poskytující předpovědi dostupnosti dopravních prostředků. Provedeným testováním byly zjištěny metriky implikující kladný přínos prediktivní analýzy ke zlepšení sdílené cyklistické dopravy ve městě Brno. Obě části aplikace ponechávají prostor zlepšením vizuálních a analytických vlastností, které mohou dále zefektivnit práci s prediktivní analýzou a uživatelskou částí aplikace.

Literatura

- [1] AKCAY, S.; AMELN, D.; VAIDYA, A.; LAKSHMANAN, B.; AHUJA, N. et al. Anomalib: A Deep Learning Library for Anomaly Detection. *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022. Dostupné z: <http://dx.doi.org/10.1109/icip46576.2022.9897283>.
- [2] BOTSMAN, R. a ROGERS, R. *What's mine is yours: the rise of collaborative consumption*. HarperBusiness, 2010. ISBN 9780061963544.
- [3] CZECH DAILY. New Regulations for E-scooters in the Czech Republic. *Czech Daily*, 2024. Dostupné z: <https://czechdaily.cz/new-regulations-for-e-scooters-in-the-czech-republic/>.
- [4] DELOITTE CANADA. *Crowdfunding: From novelty to viable capital-raising tool*. 2023. Dostupné z: <https://www.deloitte.com/ca/en/services/audit-assurance/perspectives/crowdfunding.html>.
- [5] DEMAIO, P. Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation*. University of South Florida Libraries, 2009, sv. 12, č. 4. ISSN 2375-0901. Dostupné z: <http://dx.doi.org/10.5038/2375-0901.12.4.3>.
- [6] FISHMAN, E. Bikeshare: A Review of Recent Literature. *Transport Reviews*. Informa UK Limited, 2015, sv. 36, č. 1. ISSN 1464-5327. Dostupné z: <http://dx.doi.org/10.1080/01441647.2015.1033036>.
- [7] FRANCE 24. Paris votes to ban rental e-scooters. *France 24*, 2023. Dostupné z: <https://www.france24.com/en/europe/20230402-parisians-vote-to-ban-for-hire-e-scooter-rentals-from-french-capital>.
- [8] GOUDIN, P. *The Cost of Non-Europe in the Sharing Economy: Economic, Social, and Legal Challenges and Opportunities*. European Parliamentary Research Service, 2016. Dostupné z: [https://www.europarl.europa.eu/RegData/etudes/STUD/2016/558777/EPRS_STU\(2016\)558777_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2016/558777/EPRS_STU(2016)558777_EN.pdf).
- [9] HASTIE, T.; TIBSHIRANI, R. a FRIEDMAN, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. New York: Springer, 2009. ISBN 978-0387848570. Dostupné z: <https://esl.hohoweiya.xyz/book/The%20Elements%20of%20Statistical%20Learning.pdf>.
- [10] JAMES, G.; WITTEN, D.; HASTIE, T. a TIBSHIRANI, R. *An Introduction to Statistical Learning: with Applications in R*. Springer US, 2021. Springer Texts in Statistics. ISBN 9781071614181. Dostupné z: <http://dx.doi.org/10.1007/978-1-0716-1418-1>.

- [11] KLUFOVÁ, R. *Geografické informační systémy 1*. 2024. Dostupné z: http://home.ef.jcu.cz/~klufova/GIS/GIS1/_book/.
- [12] KUMARAN, N. *Understanding Gmail's spam filters*. 2022. Dostupné z: <https://workspace.google.com/blog/identity-and-security/an-overview-of-gmails-spam-filters>.
- [13] LONGLEY, P. a GOODCHILD, M. *Geographic Information Science and Systems*. In: Elsevier, 2020. ISBN 9780081022962. Dostupné z: <http://dx.doi.org/10.1016/b978-0-08-102295-5.10557-8>.
- [14] MARTIN, E. a SHAHEEN, S. *Impacts of Car2go on Vehicle Ownership, Modal Shift, Vehicle Miles Traveled, and Greenhouse Gas Emissions: An Analysis of Five North American Cities*. Transportation Sustainability Research Center, University of California, Berkeley, 2016. Dostupné z: <https://escholarship.org/uc/item/3mc5m6pn>.
- [15] MOODY, J.; FARR, E.; PAPAGELIS, M. a KEITH, D. R. The value of car ownership and use in the United States. *Nature Sustainability*. Springer Science and Business Media LLC, 2021, sv. 4, č. 9. ISSN 2398-9629. Dostupné z: <http://dx.doi.org/10.1038/s41893-021-00731-5>.
- [16] NOVÁK, J. *Metody vizualizace dat pomocí statistického výpočetního prostředí R*. 2020. Disertační práce. VŠE Praha. Dostupné z: https://vskp.vse.cz/70622_metody_vizualizace_dat_pomoci_statistickeho_vypocetniho_prostredi_r.
- [17] *Sdílená doprava*. Dostupné z: <https://www.spolecne-udrzitelne.cz/aktuality/inspirace/sdilena-doprava/>.
- [18] SAADI, I.; WONG, M.; FAROOQ, B.; TELLER, J. a COOLS, M. An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service. *ArXiv preprint arXiv:1703.02433*, 2017. Dostupné z: <https://arxiv.org/abs/1703.02433>.
- [19] SHAH, D. *Logistic Regression: Formula, Types, and Practical Applications*. V7 Labs, 2023. Dostupné z: <https://www.v7labs.com/blog/logistic-regression>.
- [20] SHAHEEN, S. A.; GUZMAN, S. a ZHANG, H. Bike-sharing in Europe, the Americas, and Asia: Past, Present, and Future. *Transportation Research Record: Journal of the Transportation Research Board*. SAGE Publications, 2010, sv. 2143, č. 1. ISSN 2169-4052. Dostupné z: <http://dx.doi.org/10.3141/2143-20>.
- [21] SINGH, A. Why: Practical Applications of Decision Trees (Part 2). *Medium*, 2023. Dostupné z: <https://medium.com/%40diehardankush/why-practical-applications-of-decision-trees-ae09e04b2b16>.
- [22] SUNDARARAJAN, A. *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism*. The MIT Press, 2016. ISBN 9780262034573.
- [23] UPGRADED POINTS. *Airbnb vs. Hotel Cost Comparison Across the U.S.* 2024. Dostupné z: <https://upgradedpoints.com/news/airbnb-vs-hotel-cost-in-us-cities/>.

- [24] VOŠTRADOVSKÁ, A. *Sdílená ekonomika v České republice*. 2021. Diplomová práce. Univerzita Pardubice, Fakulta ekonomicko-správní, Pardubice. Vedoucí práce ING. VERONIKA LINHARTOVÁ, P. Ing. et. Dostupné z: <https://theses.cz/id/0icbxx/>.