



FAKULTA  
INFORMAČNÍCH  
TECHNologiÍ  
ČVUT V PRAZE

# Zpracování a analýza záznamů z emisních kontrol vozidel v České republice

---

Vedoucí práce: Mgr. Adam Szabó

Autor: Adam Prokop

FIT ČVUT

17. prosince 2025



## Struktura

- 1 Cíle práce
- 2 Motivace výběru tématu
- 3 Praktická část práce
  - Získání dat
  - Analýza a čištění dat
  - Detekce anomálií
  - Vytvoření modelu
  - Analýza trendů v čase
  - Vizualizace na webových stránkách
- 4 Závěr





## Cíle

- Detekce anomálií v záznamech emisních kontrol
- Vytvoření modelu pravděpodobnosti úspěchu kontroly
- Analýza agregovaných trendů v čase
- Vizualizace výsledků přehlednou formou





## Motivace

- Přispění do veřejné diskuse
- Aktualizace předchozí analýzy Ministerstva dopravy

Stupeň klasifikace dokumentů TCZ: DŮVĚRNÉ

TÜV SÜD Czech s.r.o.



Czech

4/55

### 1. ÚVOD

Na základě objednávky Ministerstva dopravy ČR zpracoval TÜV SÜD Czech s.r.o. analýzu dat o měření emisí vozidel, získaných v rámci pravidelných technických prohlídek. Zpracovateli byla dodána data exportovaná z ISTP za období 07/2017 až 08/2020. Po konzultaci se zadavatelem byl upřesněn rozsah analýzy.

**Obrázek:** Předchozí analýza emisí





## Praktická část

### Získání dat

#### 1 Lokalizace relevantních datasetů na data.gov.cz

The screenshot shows the 'Portál o datech' (Data Portal) website. The header is dark blue with the title 'Portál o datech' and a language selector 'English'. Below the header is a navigation bar with links: 'SPRÁVA DAT', 'VZDĚLÁVÁNÍ', 'ČLÁNKY', 'DATOVÁ KVALITA', 'NÁRODNÍ KATALOG OTEVŘENÝCH DAT', and 'DALŠÍ'. A search bar contains the text 'Co hledáte' and a 'Hledat' button. Below the search bar, it says 'Rozšířené vyhledávání' with a dropdown arrow. The results section shows 'Nalezeno 6241 datových sad.' and a dropdown menu for 'Název vzestupně'. Two featured datasets are listed: 'Data z měřicích přístrojů získaných při měření emisí' with a description 'Data získaná z měřicích přístrojů při provádění technické nebo emisní kontroly na stanicích STK nebo SME' and a right arrow; and 'Prohlídky vozidel STK a SME' with a description 'Otevřená data prohlídek vozidel STK a SME' and a right arrow.

Portál o datech English

SPRÁVA DAT VZDĚLÁVÁNÍ ČLÁNKY DATOVÁ KVALITA NÁRODNÍ KATALOG OTEVŘENÝCH DAT DALŠÍ

Co hledáte Hledat

Rozšířené vyhledávání

Nalezeno 6241 datových sad. Název vzestupně

Data z měřicích přístrojů získaných při měření emisí  
Data získaná z měřicích přístrojů při provádění technické nebo emisní kontroly na stanicích STK nebo SME

Prohlídky vozidel STK a SME  
Otevřená data prohlídek vozidel STK a SME

Obrázek: Portál o datech



## Praktická část

### Získání dat

## 2 Automatické stažení dat přes SPARQL endpoint

- Denní členění – tisíce záznamů
- XML formát – přes 150 GB dat

```
SELECT ?title ?downloadURL
WHERE {{
  <{parent_dataset_iri}> dcat:seriesMember ?dataset.
  ?dataset dcat:distribution ?distribution.
  ?dataset dcterms:title ?title.
  ?distribution dcat:downloadURL ?downloadURL.
  FILTER(LANG(?title) = "cs")
}}
```

### About: [Measuring instrument data from emissions measurements](#)

[Goto](#) [Sponge](#) [NotDistinct](#) [Permalink](#)

An Entity of Type : [dcat:DatasetSeries](#), within Data Space : [data.gov.cz](#) associated with source [document\(s\)](#)

Type:

Attributes	Values
<a href="#">dcterms:title</a>	Measuring instrument data from emissions measurements (en)
<a href="#">dcat:seriesMember</a>	<a href="#">Data from measuring instruments 14-08-2024</a> <a href="#">Data from measuring instruments 01-08-2024</a> <a href="#">Data from measuring instruments 30-05-2024</a> <a href="#">Data from measuring instruments 25-02-2024</a> <a href="#">Data from measuring instruments 18-11-2023</a> <a href="#">Data from measuring instruments 19-10-2023</a> <a href="#">Data from measuring instruments 05-08-2023</a> <a href="#">Data from measuring instruments 11-08-2023</a> <a href="#">Data from measuring instruments 21-03-2023</a> <a href="#">Data from measuring instruments 26-11-2024</a> <a href="#">»more»</a>

**Obrázek:** Dokumentace datasetu ve SPARQL





## Praktická část

### Analýza a čištění dat

## 1 Extrakce informací z XML do parquet formátu

```
<DatumProhlidky>2022-08-27</DatumProhlidky>
<Stanice>
  <Cislo>3237</Cislo>
</Stanice>
<CasoveUdaje>
  <Zahajeni>2022-08-27T11:58:54.0870000+02:00</Zahajeni>
  <Ukonceni>2022-08-27T12:05:37.3870000+02:00</Ukonceni>
</CasoveUdaje>
<OdpovednaOsoba>49493</OdpovednaOsoba>
<PristrojData>
  <prohlidka cisloProtokolu="CZ-003237-22-08-0921"
    datumProhlidky="2022-08-27T12:05:00">
    <mericiPristroj vyrobce="ACTIA CZ s.r.o." typ="AT505" verze="-"
      OBD="R+OBD" verzeSoftware="3.05.3"/>
  </prohlidka>
  <vozidlo>
    <VIN>WFOMXXGCDM6M38271</VIN>
    <tovazniZnacka>FORD</tovazniZnacka>
    <typVozidla>FOCUS C-MAX</typVozidla>
    <typMotoru>HWD4</typMotoru>
    <cisloMotoru>-</cisloMotoru>
    <stavTachometru>151057</stavTachometru>
    <rokVyroby>2006</rokVyroby>
    <datumPrvniRegistrace>2006-08-28</datumPrvniRegistrace>
    <palivo>BA</palivo>
  </vozidlo>
```

. . .





## Praktická část

### Analýza a čištění dat

## 2 Uložení do samostatných souborů po dnech

### ■ Nižší nároky na operační paměť

```
prohlidky_Prohlidky vozidel STK a SME za 01-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 02-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 03-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 04-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 05-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 06-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 07-01-2019.parquet  
prohlidky_Prohlidky vozidel STK a SME za 07-01-2019.parquet  
...
```

## 3 Načtení do polars

```
lazy_df = pl.scan_parquet(  
    "data/parquet/prohlidky/*.parquet",  
    schema=prohlidky_schema  
)
```

## 4 Vyřešení nekonzistencí po obsahové stránce dat

## 5 Rozdělení vozidel podle typu pohonu





## Praktická část

### Detekce anomálií

- 1 Využití statistických metod pro detekci anomálií
  - Metody nesupervizovaného učení
  - Inspirace postupy aplikovanými v analýze MDČR
- 2 Vyhledání podezřelých hodnot na základě identity záznamů

Datum	Znacka	Model	Druh	Čas 1. akcelerace [s]	Čas 2. akcelerace [s]	Čas 3. akcelerace [s]	Čas 4. akcelerace [s]
07.04.2020	FORD	GALAXY	OSOBNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
07.04.2020	AUDI	A4 (B5)	OSOBNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
07.04.2020	VOLKSWAGEN	GOLF (1J)	OSOBNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
07.04.2020	MAZDA	B	NÁKLADNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
08.04.2020	SUZUKI	GRAND VITARA	OSOBNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
08.04.2020	CITROËN	BERLINGO (G)	NÁKLADNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
08.04.2020	AVIA	A21T-FC	NÁKLADNÍ AUTOMOBIL	1,88	1,84	1,78	1,7
09.04.2020	AUDI	A6	OSOBNÍ AUTOMOBIL	1,88	1,84	1,78	1,7

Tabulka: Předchozí diskuse na x.com





## Praktická část

Model  
pravděpodobnosti  
úspěchu

- 1 Agregace informací z více datasetů na základě VIN
- 2 Natrénování několika modelů a výběr nejvýkonnějšího
- 3 Analýza důležitosti příznaků





## Praktická část

### Trendy

- 1 Analýza změn vybraných ukazatelů v čase
- 2 Predikce vývoje do budoucna



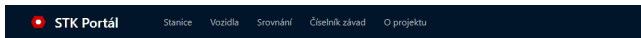


## Praktická část

### Prezentace výsledků

# 1 Zakomponování výsledků do existujících webových stránek

- Diplomová práce ing. Daniela Brotze
- Zobrazuje zajímavé infografiky související s měřením na STK
- Poskuteje predikci výsledku příští technické kontroly



## STK portál

Vítejte na STK portálu. Naleznete zde informace o stanicích technické kontroly, detaily o vozidlech v ČR, statistiky vozového parku a prohlídek na STK. Portál nabízí také srovnávač konkrétních vozidel i značek a modelů celkově. To vše na základě dat z registru silničních vozidel a záznamů o prohlídkách na STK zveřejněných Ministerstvem dopravy ČR. [Více o portálu a datech...](#)

Vozide v databázi

11 127 113

Stanice TK

386

Kontrol na STK

17 933 976

Časové pokrytí

2018–2022

Vyhledávač umožňuje najít konkrétní vozidlo podle VIN (17 znaků). Hledat lze také stanice technické kontroly podle názvu, města nebo kontaktních údajů (telefon, email).

Hledat

**Obrázek:** Aktuální verze STK portálu



## Závěr

- Automatizované zpracování a vizualizace skutečných dat
  - V reálném čase
  - Původní forma pro laiky nepoužitelná
- Identifikace podezřelých stanic
- Lepší porozumění trendům



Děkuji za pozornost