Report\_Paprnak

2025-10-10

**Projekt:** Aplikácia strojového učenia pre Big Data

**Autor:** Matej Paprnák

**Dátum:** 10.10.2025

## Súhrn

Jednoduchá SWOT analýza, študovanie právneho pozadia, kontrola bezplatných dát a dimenzovanie.

## Dokončené

* **SWOT analýza:**
  + Strengths (silné stránky):
    - Akademický prínos
    - Možnosť open-source rozvoja
  + Weaknesses (slabé stránky):
    - Právna neistota
    - Časová náročnosť spracovania dát
    - Vysoké nároky na výpočtový výkon
    - Potenciálna zaujatosť (málo dialektov v rádiách)
  + Opportunities (príležitosti):
    - Zlepšenie slovenských systémov
    - Spolupráca s médiami
    - Grantové financovanie
    - Otvorený slovenský dataset (pod podmienkou anonymizácie)
  + Threats (hrozby)
    - Porušenie autorských práv
    - Porušenie GDPR
    - Finančné limity
    - Rýchly technologický vývoj
* **Študovanie právneho pozadia:** Keďže na projekte pracujeme cez výskumnú inštitúciu, vzťahuje sa na nás TDM výnimka, ktorá umožňuje robiť kópie diel bez povolenia nositeľov za určitých podmienok. Tieto podmienky sú spísané v nasledujúcom dokemente:

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32019L0790>

Po pôvodnom čítaní som dospel k záveru, že všetky podmienky spĺňame, no je potrebné si overiť či streamy staníc, ktoré budeme používať sú verejne dostupné bez DRM. Okrem autorského práva je však potrebné brať na vedomie aj ochranu osobných údajov (GDPR). Hlasové záznamy sa totiž dajú použiť na identifikovanie rečníka, a teda sa považujú za obsahujúce osobné údaje, na ktoré zber potrebujeme legálny základ. Najjednoduchšia cesta k jeho získaniu, ktorá mi napadá je písomný súhlas rečníkov. Alternatívne môžeme dáta anonymizovať, a teda upraviť do bodu, v ktorom idetifikovať rečníka z nadobudnutého hlasového záznamu nebude možné. Týmto však riskujeme problémy pri eventuálnom trénovaní modelu na týchto dátach. Podobné a pravdepodobne závažnejšie problémy môžu nastať pri práci s hudobnou tvorbou, ktorá tvorí drvivú väčšinu rádiového vysielania na väčšine staníc, a taktiež komerčných prestávkach.

* **Common voice:** Na tejto stránke existujú bezplatné kolekcie dát vo viacerých jazykoch, vrátane toho slovenského, vďaka čomu by sa dala využiť v úvodných fázach projektu bez rizika porušenia akýchkoľvek práv a zákonov. Spolu sa tam nachádza 472 nahratých hodín, uložených v 30 balíčkoch o rôznych veľkostiach. Dáta sú exkluzívne vo formáte MP3.
* **Dimenzovanie:** Najpopulárejnejšie slovenské rádiové stanice streamujú v kvalite 128kb. To znamená, že na uloženie jedného 24-hodinového záznamu z jednej takejto stanice budeme potrebovať

kb = 128 \* 60 \* 60 \* 24  
kB = kb / 8  
MB = kB / 1024  
GB = MB / 1024  
paste0(GB, " GB.")

## [1] "1.318359375 GB."

## Použitie umelej inteligencie

* **Prompt:** I am supposed to create a report of my progress regarding a certain project I’m part of. How does such a document look like?
* **Výstup:** Good question — a **project progress report** is one of those documents that sounds formal but is actually quite straightforward once you know what it’s meant to do.

It’s basically a *snapshot* of:

* what you’ve done,
* what’s being worked on,
* what’s next,
* and whether anything’s blocking progress.

Here’s what a typical structure looks like 👇

## 🧱 **Typical structure of a project progress report**

### 1. **Header / Title**

Include:

* Project name
* Your name and role
* Date (and reporting period, e.g. *Week 3 – Oct 1–5, 2025*)
* Possibly who the report is for (e.g. *Supervisor: Dr. Novak*)

### 2. **Summary / Overview**

A short paragraph (3–5 sentences) summarizing the project’s current state. Example:

The project “AI Traffic Flow Simulation” is progressing according to plan. Data preprocessing and baseline model training have been completed. The focus this week has been on optimizing the reinforcement learning environment. Minor delays occurred due to hardware availability, but overall progress remains on track.

### 3. **Completed Work**

A clear, concise list of what you’ve done since the last report. You can format it as bullets or a short table.

**Example:**

* Collected and cleaned sensor data for three intersections.
* Set up data pipeline in Python.
* Trained initial prediction model (accuracy: 82%).
* Fixed CSV import bug affecting timestamps.

### 4. **Current Work / In Progress**

Explain what you’re currently doing and how it contributes to project goals.

Implementing a custom reward function for traffic optimization using reinforcement learning. Evaluating simulation results with new parameters.

### 5. **Next Steps**

List what’s planned for the next phase or reporting period.

* Fine-tune the RL model parameters.
* Add visualization dashboard for simulation output.
* Prepare presentation slides for progress meeting.

### 6. **Issues / Risks / Challenges**

Briefly note any obstacles or things that could affect the timeline.

Model training takes longer than expected due to GPU queue limits. Considering using smaller datasets for initial tests.

### 7. **Additional Notes (optional)**

* Feedback or requests for support.
* Interesting findings or unexpected results.
* Dependencies on other team members.

## 🧾 **Example layout**

Project: Data Visualization Tool for Iris Dataset  
Author: Matej Paprnák  
Date: October 5, 2025  
Reporting period: Week 4  
  
---  
  
### Summary  
The core visualization module has been implemented. It now supports basic scatter plots for the Iris dataset and interactive selection of variables. Integration with the UI framework is ongoing.  
  
### Completed Work  
- Finished setup of R Markdown environment.  
- Implemented scatter plot generation.  
- Added function for data filtering by species.  
- Uploaded initial code and documentation to GitHub.  
  
### In Progress  
- Connecting plots to UI controls.  
- Testing export to PDF functionality.  
  
### Next Steps  
- Finalize UI linking.  
- Write user documentation.  
- Conduct test with example dataset.  
  
### Issues  
Minor issues with R Markdown figure placement in PDF output.

Would you like me to format a **progress report template** (in Markdown or Word/PDF style) that you can just fill in with your info and upload to GitHub?