D21 – Model pre predikciu verdiktov (SK)

Projekt financovaný Európskou úniou Next GenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00049.

Financované Európskou úniou NextGenerationEU
PLÁN OBNOVY
ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU
VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA
VÝSKUMNÁ AGENTÚRA

ÚVOD

Projekt „**Automatizácia analýzy právnych textov na základe strojového učenia**“ (ďalej len „**ALTAML**“) predstavuje kľúčovú iniciatívu zameranú na integráciu inovatívnych prístupov v oblasti spracovania údajov a následnej analýzy, so zameraní na oblasť práva, ktorá zahŕňa aj právo informačných a komunikačných technológií. Cieľom tohto projektu je vyvinúť a overiť účinné metódy automatizovanej analýzy právnych textov pomocou techník strojového učenia. To zahŕňa najmä vývoj nástrojov, ktoré uľahčujú spracovanie a analýzu veľkých objemov údajov vo forme rôznych právnych dokumentov, ako aj extrakciu relevantných informácií (atribútov) z takýchto dokumentov vrátane identifikácie kľúčových pojmov, odkazov na právne predpisy a ďalších atribútov.

Cieľom projektu ALTAML je teda prispieť k efektívnejšiemu prístupu k právnym informáciám a k zrýchleniu právnych procesov, čím sa zabezpečí vyššia miera právnej istoty a zlepší sa dostupnosť právnych textov, výsledkov ich analýzy a relevantných právnych informácií pre odbornú aj širokú verejnosť.

V rámci pracovného balíka KPB4 bol vypracovaný výstup Model pre predikciu verdiktov. Navrhnutý systém predstavuje hybridný právny analytický pipeline pre automatizovanú klasifikáciu výsledku trestného konania zo slovenských súdnych rozhodnutí a následnú extrakciu uložených trestov. Riešenie kombinuje deterministické pravidlá založené na právnych ustanoveniach, sémantické vektorové reprezentácie dokumentov a veľký jazykový model lokálne nasadený prostredníctvom nástroja Ollama.

Systém využíva viacvrstvovú architektúru typu Tier 0 – Tier 2, kde jednotlivé vrstvy zabezpečujú postupné rozhodovanie s rastúcou mierou sémantického spracovania. Primárnym cieľom systému je identifikovať, či bol obžalovaný oslobodený alebo odsúdený, a v prípade odsúdenia automaticky extrahovať uložené tresty vrátane ich typu, miery, podmienok a detailov výkonu trestu.

Na zvýšenie presnosti rozhodovania systém využíva Retrieval-Augmented Generation (RAG) prístup založený na vyhľadávaní historicky podobných prípadov pomocou embeddingov. Každé rozhodnutie je reprezentované vektorovou reprezentáciou vytvorenou pomocou embedding modelu BGE-M3, pričom podobnosť prípadov je vyhodnocovaná prostredníctvom kosínovej podobnosti medzi embeddingami jednotlivých dokumentov. Najrelevantnejšie historické prípady sú následne použité ako few-shot príklady pre veľký jazykový model.

Finálne rozhodovanie o oslobodení a extrakcii trestov vykonáva lokálne nasadený jazykový model LLaMA 3.3, ktorý je riadený striktne definovanými promptami

formulovanými v slovenskom jazyku. Výstupy modelu sú deterministicky obmedzené na JSON štruktúry, čím je zabezpečená konzistentnosť a strojová spracovateľnosť výsledkov. Stručnejší popis nájdete nižšie.

Hybridný systém predikcie výsledku trestného konania a extrakcie trestov zo slovenských súdnych rozhodnutí

Systém predstavuje softvérový modul implementovaný v jazyku Python, navrhnutý na automatickú analýzu slovenských trestných rozhodnutí. Pipeline kombinuje pravidlové právne rozhodovanie, sémantické vyhľadávanie podobných prípadov a veľký jazykový model na dosiahnutie presnej klasifikácie rozhodnutí a extrakcie uložených trestov.

Funkcionalita systému

Hlavná funkcionalita systému pozostáva z nasledujúcich krokov:

1. Príprava vstupných dát

Systém pracuje s predspracovanými štruktúrovanými rozhodnutiami, ktoré obsahujú:

- skutkovú vetu,
- identifikované trestné činy,
- poľahčujúce a priťažujúce okolnosti,
- identifikované dôkazy,
- odkazované právne predpisy,
- metadáta o rozhodnutí.

Vstupné texty sú agregované do jednotnej kontextovej reprezentácie, ktorá je následne použitá na embeddingy aj inferenciu jazykového modelu.

2. Sémantické reprezentácie dokumentov

Každé rozhodnutie je reprezentované embeddingom vytvoreným pomocou modelu BGE-M3. Embedding zahŕňa:

- skutkový opis,
- identifikované trestné činy,
- poľahčujúce okolnosti,
- priťažujúce okolnosti,
- identifikované dôkazy.

Embeddingy sú ukladané do lokálnej databázy v JSON formáte, čím sa minimalizuje potreba opakovaných výpočtov pri ďalšom spracovaní.

3. Viacvrstvová klasifikácia rozhodnutí

3.1. Tier 0 – Detekcia zastavenia konania

Prvá vrstva systému identifikuje prípady, ktoré boli zastavené alebo prerušené na základe explicitných paragrafových odkazov v rozhodnutí. Takéto prípady sú automaticky vyradené z ďalšieho spracovania.

3.2. Tier 1 – Pravidlová klasifikácia oslobodenia

Druhá vrstva využíva deterministické právne pravidlá založené na identifikovaných paragrafových fragmentoch.

Systém analyzuje výskyt:

- paragrafov signalizujúcich oslobodenie,
- paragrafov signalizujúcich odsúdenie,
- dominantných odsudzujúcich ustanovení.

Na základe týchto pravidiel systém rozhodne:

- oslobodenie = true,
- oslobodenie = false,
- alebo neurčitý výsledok, ktorý postupuje do ďalšej vrstvy.

3.3. Tier 2 – Hybridné rozhodovanie pomocou RAG a LLM

Ak pravidlová vrstva nedokáže spoľahlivo rozhodnúť, systém aktivuje hybridný režim založený na Retrieval-Augmented Generation (RAG).

Pre aktuálny prípad sa:

- vypočíta embedding rozhodnutia,
- vyhľadajú najpodobnejšie historické prípady,
- vyberú few-shot príklady s najvyššou kosínovou podobnosťou.

Tieto prípady sú následne vložené do promptu veľkého jazykového modelu, ktorý vykoná finálne rozhodnutie o oslobodení alebo odsúdení.

4. Extrakcia uložených trestov

Ak je rozhodnutie klasifikované ako odsudzujúce, systém vykoná automatickú extrakciu uložených trestov.

Jazykový model identifikuje:

- typ trestu,
- mieru trestu,
- podmienku výkonu trestu,
- kategóriu trestu,
- doplňujúce detaily výkonu.

5. Retrieval-Augmented Generation (RAG)

Kľúčovou súčasťou systému je mechanizmus vyhľadávania podobných prípadov. Pre každý analyzovaný dokument sa:

- vypočíta embedding,
- porovná sa s embeddingami historických rozhodnutí,
- vyberú sa najrelevantnejšie prípady pomocou kosínovej podobnosti.

Tieto prípady sú následne použité ako few-shot príklady pri inferencii veľkého jazykového modelu, čím sa zvyšuje stabilita a právna konzistentnosť výstupov.

6. Prompt engineering a deterministické výstupy

Systém využíva striktne definované slovenské prompty navrhnuté pre:

- klasifikáciu oslobodenia,
- extrakciu trestov,
- validáciu štruktúry výstupov.

Model je explicitne obmedzený:

- na výstup vo formáte JSON,
- na použitie definovaných typov trestov,
- na konzistentné používanie právnej terminológie.

Inferencia prebieha s nulovou teplotou (temperature = 0.0), čím sa minimalizuje variabilita generovaných odpovedí.

Výstup systému

Výstupom systému je JSON štruktúra, ktorá pre každý identifikátor rozhodnutia obsahuje:

- informáciu o oslobodení alebo odsúdení,
- zoznam extrahovaných trestov,
- mieru a podmienky trestu,
- informáciu o preskočených prípadoch.

Výsledky sú ukladané priebežne počas spracovania, čím je zabezpečená odolnosť pipeline voči prerušeniu výpočtu.

Požiadavky a technológie

Systém je implementovaný v jazyku Python a využíva nasledovné technológie:

- lokálny LLM (LLaMA 3.3) pre právnu inferenciu,
- embedding model BGE-M3 pre sémantické reprezentácie,
- kosínovú podobnosť na vyhľadávanie podobných prípadov,
- Retrieval-Augmented Generation (RAG),
- JSON databázu embeddingov pre cacheovanie reprezentácií,
- Ollama API pre lokálnu inferenciu modelov.

Odporúča sa RAM 32 GB a viac. GPU akcelerácia (CUDA) výrazne zrýchľuje výpočet embeddingov aj inferenciu jazykového modelu pri spracovaní rozsiahlych právnych korpusov.

Zdrojové súbory

Zdrojový kód projektu a dátový súbor sú udržiavané v repozitári a sú dostupné na adrese: https://gitlab.science.upjs.sk/altaml/verdict-prediction