

# LoL Intra League Predictor

Ladislav Pavláček

ČVUT–FIT

pavila5@fit.cvut.cz

31. prosince 2025

„Pro zvýšení čitelnosti kódu a dodržení standardů PEP 8 byly dokumentační řetězce (docstrings) vygenerovány pomocí generativního modelu AI Gemini [2]. Všechny vygenerované texty byly následně manuálně revidovány.“

## 1 Motivace

Hlavní motivací pro výběr tématu predikce výsledků zápasů v League of Legends byla snaha aplikovat metody strojového učení na oblast profesionálního e-sportu. Toto téma mi umožnilo propojit teoretické znalosti z předmětu BI-ML1 a BI-PYT s osobním zájmem o e-sportové prostředí. Cílem bylo vytvořit funkční projekt, který pokrývá celý proces od vlastního sběru dat až po jejich analýzu, a pracovat tak na problému, který je pro mě osobně zajímavý a motivující.

## 2 Úvod

Projekt se zaměřuje na predikci výsledků profesionálních zápasů v League of Legends v rámci nejvyšších regionálních lig. Práce pokrývá kompletní proces zpracování dat a tvorby modelu: od automatizovaného sběru dat (scraping), přes jejich analýzu a preprocessing, až po trénování a testování prediktivních modelů. Součástí výsledného řešení je také uživatelské rozhraní (GUI). Následující sekce se věnují jednotlivým fázím vývoje, přičemž detailní technická dokumentace a instrukce ke spuštění jsou k dispozici v souboru README v přiloženém repozitárii.

## 3 Vstupní data

Pro získání potřebných informací byly využity dva primární zdroje: server Oracle's Elixir pro detailní statistiky týmů a GOL.gg pro výsledky jednotlivých zápasů.

Získávání dat z Oracle's Elixir bylo přímočaré díky dostupnosti ročních statistik v uceleném formátu CSV. Naopak pro sběr dat o zápasech z GOL.gg bylo nutné implementovat vlastní automatizovaný scraper. Ten je navržen jako modulární

nástroj, kde uživateli stačí specifikovat názvy požadovaných turnajů (viz README) a skript automaticky zajistí kompletní stažení dat.

Vzhledem k tomu, že statistiky týmů tvořily rozsáhlý a komplexní soubor obsahující veškerá herní data, byla kritickou částí vývoje implementace logiky pro filtraci a konkretizaci relevantních parametrů. Jednotlivé příznaky jsou detailněji popsány v README.

Kvalita dat je zajištěna automatizovaným testováním vypočtených statistik s referenčními daty ze serveru Oracle's Elixir.

## 4 Algoritmus pro slučování dat

Klíčovou výzvou projektu bylo navržení algoritmu pro sloučení dat o zápasech se statistikami týmů tak, aby výsledný dataset reflektoval aktuální herní formu. Data byla segmentována podle herních období (splitů), přičemž výpočet statistik probíhá dynamicky. Algoritmus využívá tři scénáře pro zajištění relevantní historie dat:

- **Stabilní fáze:** Pokud tým v aktuálním splitu odehrál více než 5 zápasů, počítají se statistiky (průměry) pouze z těchto odehraných her v rámci daného splitu.
- **Zahájení sezóny:** U úplně prvního zápasu splitu se pro predikci využívají kompletní statistiky z předchozího odehraného splitu.
- **Přechodná fáze:** Pokud má tým odehráno méně než 5 zápasů, kombinují se aktuální data s redukovaným vzorkem z minulého splitu (limitovaným na 5 her). Tím je zajištěna dostatečná datová báze při zachování vyšší váhy aktuálních výsledků.



Obrázek 1: Team-Match Linking Logic Y

## 5 Trénování a validace dat

Z důvodu dynamického vývoje hry (změny patchů a soupisek) byla jako validační množina zvolena nejnovější data (poslední 2 měsíce). Pro zvýšení přesnosti byla zavedena váha vzorků podle data konání zápasu na spojité škále 0 (nejstarší data) až 1 (nejnovější data) - trénovací množina. Po optimalizaci hyperparametrů je finální model přetrenován na kompletním datasetu - nutné pro lepší generalizaci.

Vzhledem k charakteru domény je ideálním testováním dlouhodobé nasazení v reálném čase. Model vyžaduje pravidelné měsíční přetrenování na nejnovějších datech, aby reflektoval aktuální herní metriky a složení týmů. Pro objektivní zhodnocení úspěšnosti modelu v reálném provozu je nezbytná alespoň dvouměsíční fáze monitorování a kontinuální vyhodnocování predikcí nad nově vznikajícími daty.

## 6 Budoucnost projektu

Projekt nabízí široký prostor pro další rozvoj a optimalizaci. Mezi hlavní oblasti potenciálního zlepšení patří:

- **Pokročilé metriky:** Implementace specifických ukazatelů jako EGR (*Early Game Rate*) a MLR (*Mid/Late Game Rate*) pomocí lineární regrese vybraných příznaků pro lepší zachycení dynamiky jednotlivých fází hry.
- **Rozšíření datasetu:** Navýšení objemu trénovacích dat a integrace dalších e-sportových titulů (např. CS:GO), což by vyžadovalo úpravu scrapperů a redefinici herních parametrů.
- **Technologický upgrade:** Experimentování s metodami hlubokého učení (*Deep Learning*), které by mohlo lépe zachytit nelineární vztahy v herních datech.
- **Optimalizace rozhraní:** Další vývoj grafického uživatelského rozhraní (GUI) pro intuitivnější zobrazení predikcí a statistik v reálném čase.

Dlouhodobé sledování úspěšnosti modelu v reálném provozu by navíc poskytlo cennou zpětnou vazbu pro kontinuální ladění přesnosti a robustnosti celého systému.

## 7 Závěr

Práce na tomto projektu pro mě byla velmi obohacující a umožnila mi prohloubit znalosti v oblasti

analýzy dat a strojového učení. Během vývoje jsem narazil na řadu technických výzev, z nichž většinu se podařilo úspěšně vyřešit, zejména v oblasti automatizovaného sběru a komplexního zpracování dat a některé jsou mi záhadou do teď.

Vzhledem k časové náročnosti projektu vnímám veliký prostor pro zlepšení v oblasti ladění hyperparametrů a hlubší evaluace modelů. Hlavní důraz byl však kladen na robustní implementaci celého workflow – od získání surových dat až po funkční predikční systém s uživatelským rozhraním. Celkově hodnotím práci na projektu velmi pozitivně, neboť ve mně probudila hlubší zájem o problematiku analýzy e-sportových dat a aplikaci Pythonu v reálných úlohách.

## Reference

- [1] Games of Legends. Tournament list. online, 2025. Dostupné z: <https://gol.gg/tournament/list/>. [cit. 2025-12-31].
- [2] Google. Generování dokumentačních řetězců (docstrings) pro python kód. AI model: Gemini 1.5 Flash, 2025. Verze z 2025-12-31. Dostupné z: <https://gemini.google.com/>. [cit. 2025-12-31].
- [3] Oracle's Elixir. Team stats by tournament. online, 2025. Dostupné z: <https://oracleselixir.com/stats/teams/byTournament>. [cit. 2025-12-31].