

Projektni seminar

Razvoj eksperimentalnega sistema za simulacijo sodelovanja LLM-jev na napovednih trgih

Projektna dokumentacija

Avtor: Marjan Meglen

Mentor: dr. Aleksandar Tošić

Somentor: dr. Niki Hrovatin

UP FAMNIT, Koper

Maj 2026

Kazalo

1	Opis problema	2
2	Sistemska analiza in načrtovanje	2
2.1	Metodologija	2
2.2	Funkcionalna specifikacija	3
2.3	Modeliranje	3
3	Implementacija	4
3.1	Izbrana orodja in tehnologije	4
3.2	Pristopi	4
4	Delovni proces in časovnica	4
5	Zaključek	5

1 Opis problema

Kratka predstavitev problema

Napovedni trgi združujejo informacije več udeležencev in jih pretvarjajo v ceno oziroma verjetnost dogodka. Pri uporabi velikih jezikovnih modelov na takih trgih se pojavi vprašanje, kako se model odziva na različne vrste informacij: tržno vprašanje, trenutno tržno verjetnost, novice in opis drugih udeležencev trga. Ročno testiranje posameznih pozivov ni dovolj ponovljivo, saj je težko zagotoviti enake pogoje čez več trgov, datumov in eksperimentalnih nastavitvev.

Projekt obravnava ta problem z razvojem eksperimentalnega sistema, ki lahko za isti trg in isti datum nadzorovano spreminja informacije v pozivu ter zabeleži, kako se spremeni napoved modela `p_yes`. Sistem tako omogoča primerljivo izvedbo večjega števila LLM napovedi in pripravo rezultatov za statistično analizo.

Cilj projekta

Cilj projekta je izdelati delujoč sistem, ki podpira celoten eksperimentalni tok:

- zbiranje in pripravo podatkov o napovednih trgih, dnevnih tržnih verjetnostih in novicah,
- sestavljanje standardiziranih LLM pozivov glede na izbrano konfiguracijo,
- izvajanje napovedi z LLM modelom ali testnim mock odjemalcem,
- shranjevanje uporabljenih pozivov, odgovorov modela in izračunanih verjetnosti,
- pripravo tabel in grafov za analizo vpliva posameznih informacijskih pogojev.

Ciljna publika

Sistem je namenjen predvsem uporabnikom, ki želijo izvajati nadzorovane eksperimente z LLM modeli na napovednih trgih. Tipični uporabniki so študenti, raziskovalci in razvijalci, ki potrebujejo ponovljiv eksperimentalni okvir, ne samo enkratnega pogovora z modelom.

2 Sistemska analiza in načrtovanje

2.1 Metodologija

Projekt je bil zasnovan kot modularen eksperimentalni sistem. Namesto ene monolitne skripte so naloge razdeljene na podatkovni pipeline, izvajalnik eksperimentov in analitični del. Takšna delitev omogoča, da se posamezen del sistema spremeni brez posega v celoten projekt. Na primer, mogoče je dodati nov vir novic, zamenjati LLM model ali razširiti analizo rezultatov.

Eksperimentalna zasnova temelji na faktorskem spreminjanju treh pogojev:

- ali model vidi trenutne tržne verjetnosti,
- ali model vidi novice za trenutni dan,
- ali je v pozivu navedeno, da so drugi udeleženci LLM agenti, ljudje ali pa ta informacija ni podana.

Kombinacija teh pogojev tvori 12 konfiguracij. Za vsak trg se konfiguracije izvajajo čez 20 dnevni opazovanj, kar omogoča primerjavo napovedi v nadzorovanih pogojih.

2.2 Funkcionalna specifikacija

Sistem mora podpirati naslednje funkcionalnosti:

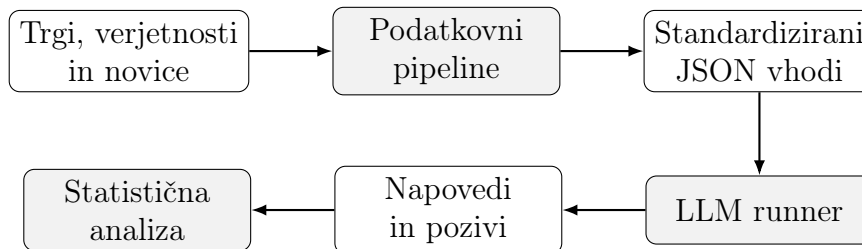
Funkcionalnost	Opis
Priprava trgov	Sistem hrani identifikator, vprašanje, datum razrešitve in status razrešitve trga.
Priprava dnevnih podatkov	Za vsak trg in datum se hrani verjetnost YES in NO.
Priprava novic	Novice so povezane s trgom in datumom ter vključujejo naslov, besedilo in metapodatke o relevantnosti.
Konfiguriranje pozivov	Uporabnik izbere konfiguracijo, ki določa vključitev tržnih verjetnosti, novic in opisa drugih udeležencev.
Izvedba LLM napovedi	Sistem za vsak dan zgradi poziv, pokliče izbranega odjemalca in zahteva strukturiran JSON odgovor.
Validacija odgovorov	Odgovor mora vsebovati kratko utemeljitev in številsko vrednost <code>probability_yes</code> med 0 in 1.
Shranjevanje rezultatov	Rezultati se shranijo po zagonih v datoteko <code>predictions.json</code> .
Analiza	Rezultati se združijo v tabele, grafe in primerjalne statistične teste.

2.3 Modeliranje

Osnovni podatkovni model vsebuje pet glavnih vhodnih entitet:

- **Market**: identifikator trga, vprašanje in datum razrešitve,
- **DailyDataRow**: dnevni podatek o verjetnosti YES in NO,
- **NewsItem**: novica, povezana s trgom in datumom,
- **Instruction**: osnovno navodilo, ki ga dobi model,
- **Configuration**: eksperimentalni pogoji, ki določajo vsebino poziva.

Izhodna entiteta je **Prediction**. Vsebuje trg, konfiguracijo, datum, utemeljitev modela, `p_yes`, `p_no`, celoten uporabljeni poziv in celoten pogovor z modelom.



3 Implementacija

3.1 Izbrana orodja in tehnologije

Projekt je implementiran v programskem jeziku Python. Podatki so shranjeni v JSON datotekah, ker je format pregleden, enostaven za validacijo in primeren za povezovanje med posameznimi moduli. Za analizo rezultatov je uporabljen Jupyter zvezek, izhodi pa so shranjeni kot CSV tabele, LaTeX tabele in slike.

Za LLM klice sistem podpira dva načina:

- `mock` odjemalec, ki deterministično vrne 0.5 in je namenjen testiranju cevovoda,
- `ollama` odjemalec, ki uporablja Ollama-kompatibilen API in strukturiran izhodni format.

Pri podatkovnem delu so uporabljeni tudi `embedding`i za izbor relevantnih člankov in dodatno semantično filtriranje novic.

3.2 Pristopi

Implementacija sledi načelu ločitve odgovornosti. Podatkovni del pripravi vhodne datoteke `markets.json`, `daily_data.json` in `news.json`. Eksperimentalni del teh podatkov ne zbira, temveč jih samo naloži, validira in uporabi za gradnjo pozivov. Analitični del pa iz rezultatov pripravi povzetke in primerjave.

Pomemben pristop je tudi sprotno shranjevanje rezultatov. Po vsaki dnevni napovedi se datoteka `predictions.json` posodobi, zato izguba enega kasnejšega LLM klica ne izbriše že pridobljenih rezultatov. Pri napakah se shrani tudi diagnostična datoteka z vhodnim pozivom in neobdelanim odgovorom.

4 Delovni proces in časovnica

Projekt je nastajal iterativno. Najprej je bila pripravljena osnovna struktura eksperimenta in podatkovni model. Nato je bil implementiran LLM runner z validacijo vhodnih podatkov

in izhodnih odgovorov. V naslednji fazi je bil dodan podatkovni pipeline za trge, tržne verjetnosti in novice. Na koncu so bili izvedeni eksperimenti ter pripravljene tabele in grafi za predstavitev rezultatov.

Obdobje	Glavno delo
Tedni 1–3	Zasnova sistema, ciljev, podatkovnih formatov in eksperimentalnih pogojev.
Tedni 4–6	Implementacija LLM runnerja, mock odjemalca, validacije in shranjevanja rezultatov.
Tedni 7–10	Razvoj podatkovnega pipelinea za trge, dnevne verjetnosti, novice in končni izvoz.
Tedni 11–12	Priprava analitičnega zvezka, grafov, tabel in preliminarne poročila.
Teden 13	Zaključni zagon eksperimentov, povzetek ugotovitev in priprava predstavitve.

5 Zaključek

Končni rezultat je delujoč eksperimentalni sistem, ki poveže pripravo podatkov, nadzorovano sestavljanje pozivov, izvajanje LLM napovedi in analizo rezultatov. Sistem omogoča ponovljive primerjave med različnimi informacijskimi pogoji. V izvedenem obsegu vključuje 6 trgov, 12 konfiguracij in 20 dnevnih napovedi na kombinacijo trga in konfiguracije, skupaj 1440 napovedi.

Najpomembnejša ugotovitev pri dosedanji analizi je, da ima prikaz trenutnih tržnih verjetnosti najizrazitejši in najbolj konsistenten vpliv na napovedi modela. Učinek novic in opis drugih udeležencev sta bolj odvisna od posameznega trga in konfiguracije.