Predlog diplomskog/master rada

Dužina i format teksta predloga projekta nisu bitni, važno je samo da precizno definišete sledeće elemente:

1. **Definicija problema/cilja projekta.** Pravljenje chatbota koji kao ulaz prima pitanje, traži najsličnije pitanje u bazi i na njega vraća odgovor (koji se isto nalazi u bazi), sentiment tog odgovora i bolduje ključne reči u odgovoru.
2. **Motivacija problema rešavanog u projektu.** Ovaj chatbot bi mogao da se koristi na raznim sajtovima i pomogao bi ljudima da brže dođu do odgovora na pitanje, odmah bi znali da li je odgovor pozitivan ili negativan (ili možda neutralan) i dobili bi i ključne reči tog odgovora kako im neki bitan deo ne bi promakao.
3. **Relevantna literatura (minimum 3 rada)**

Dajtekratak pregled onoga što se zna u o datom problemu u literaturi. Trebalo bi navesti barem 3 relevantne reference (naučni radovi). Idealno, pronađite radove koji rešavaju isti problem (isti ulazi/izlazi sistema) i identifikujte state-of-the-art u literaturi (ovo ćete svakako morati uraditi prilikom pisanja teksta diplomskog).

Za svaki od navedenih radova napisati jedan pasus koji obuhvata:

1. Šta je bio zadatak rada (šta je predviđano, kakvi su bili ciljevi i slično)
2. Koja se metodologija koristila (npr. primena Naivnog Bajesa, primena regresije, ...)
3. Kakav je bio skup podataka (naznaka kavi su atributi, referenca na rad u kome je skup podataka objavljen)
4. Kako je evaluirano rešenje (unakrsna validacija/podela na trening/test skup, ... i koja se mera koristila – accuracy/R2...)
5. Koji su najvažniji rezultati u rečenici do dve (neki interesantni zaključci tipa „pokazano je da se metodom x može dobro predvideti y, ali ne i z“ / „dostignuta je tačnost od x%“ / „metoda x je za ovu primenu bolja od metode y“)
6. (Opciono) šta po Vama nedostaje ovom rešenju? (“Autori nisu uzeli u obzir…”/ “bilo bi lepo dopuniti sa”/”Radi dobro na x, ali pitanje je kako radi na y”)

U ovoj fazi nije potrebno dati pun pregled relevantne literature koju ćete koristiti u diplomskom, ali je potrebno da ste svesni o tome šta su drugi istraživači radili na sličnu temu. Na primer, proverite da li projekat koji želite da uradite nije neko već rešavao na potpuno isti način kao što ste planirali.

**Prilikom pretrage literature se trudite da radovi budu što skorašnjiji** (izdati u poslednjih 5 godina). Dobar sajt za pretragu je *google scholar* [*https://scholar.google.com/*](https://scholar.google.com/)*.* Ako rad nije javno dostupan, možete ga potražiti putem <http://sci-hub.tw/>. Za knjige je dobar sajt <https://libgen.io/>.

1. **Skup podataka.** Opišite skup/skupove podataka koje planirate da koristite u projektu.

Postojeći (javno dostupan) skup podataka:

* Stavite referencu na rad u kome skup podataka konstruisan i link gde se skup podataka može skinuti
* Naglasite koji atribut predstavlja ciljno obeležje i šta je njegov sadržaj (klasifikacija – koje klase postoje i koja je raspodela podataka po klasama, regresija – koji je opseg u pitanju)
* Koji atributi/grupe atributa postoje u skupu podataka (na osnovu čega predviđate)

1. **Metodologija**.Kratak Chatbot će imati 4 modela: high-recall, high-precision, model za klasifikaciju sentimenta i model za ključne reči. Za pronalaženje N najsličnijih pitanja koristiću kosinusno, menhetn ili euklidsko rastojanje, zavisi koje se najbolje pokaže, a za pronalaženje najsličnijeg pitanja koristiću LSTM. Što se tiče određivanja sentimenta koristiću NN, RF, SVM i KNN i izabrati najbolji. Za ključne reči moram još pogledati kako se radi.
2. **Metod evaluacije.** Definišite kako planirate da evaluirate razvijeni sistem. Ovo je takođe najbolje definisati na osnovu toga što ste pročitali u literaturi. Obavezno definišite:
   1. **Mere performanse** (npr. i/ili RMSE za regresiju; tačnost, preciznost, odziv, ili f-mera za klasifikaciju; heterogenost + kvalitativna analiza za klasterovanje i slično).
   2. **Postupak evaluacije (eksperiment)** – npr. unakrsna validacija, podela na trening/validacioni/test skup (definisanti odnos).
3. **(Opciono) Softver**