Predmet: Veštačka inteligencija

Projektni zadatak: Praćenje aktivnosti studenata na društvenim mrežama

Profesor: Studenti:

Leonid Stoimenov Teodora Djordjević

Lazar Kadić

1. Uvod

Društvene mreže predstavljaju veoma bitnu stavku u svakodnevnom životu mladih, ponekad i sa većim prioritetom nego obaveze iz "realnog" sveta. Upravo zbog tolikog značaja i razvoja društvenih mreža, moguće je i dosta toga naučiti na njima, kako o pojedincu tako i o posmatranoj ciljnoj grupi. U našem slučaju, ciljna grupa su studenti, a konkretna društvena mreža koju razmatramo je Reddit. Reddit predstavlja platformu sličnu Twitter-u ili Facebook-u, pri čemu je veći akcenat stavljen na jasnu podelu tema (subreddit-a) o kojima se diskutuje, iznosi svoje mišljenje i stavovi, ali i zatraži pomoć za konkretan problem, što može biti od velike koristi upravo studentima. Analizom takvih objava, i generalno analizom objava pojedinaca, možemo proceniti njihovo zadovoljstvo i osećanja o određenoj temi, pa samim tim korigovati dalji pristup kako bi poboljšali subjektivni osećaj o konkretnim temama, ali i celokupno radno okruženje.

1. Projekat

Kako smo ranije pomenuli, za realizaciju ovog projekta koristiće se platforma Reddit. Reddit je web sajt u vidu foruma zasnovan na principu postavljanja, komentarisanja i ocenjivanja web sadržaja. Korisnici ili zajednice na sajtu postavljaju sadržaj u vidu teksta ili nekog drugog formata. Takav sadržaj je podeljen u subreddit-e koji pokrivaju neku odredjenu temu. Kako se na ovakvom web sajtu nalazi pregršt podataka koji mogu biti prikupljani, analizirani i obrađivani, stvorila se i potreba za aplikacijama koje će to omogućiti. Danas, uz pomoć veštačke inteligencije, a na osnovu podataka, mogu se izvoditi i određeni zaključci.

Upravo zbog prethodno navedenog, za obradu ove teme implementirana je aplikacija u Python-u koja će koristiti određene biblioteke i doći do određenih zaključaka.

1. Zadatak aplikacije

Zadatak aplikacije je da na osnovu navedene zajednice, u nasem slucaju je to zajednica studenata pod nazivom “College”, koja pise o temama koje su aktuelne među studentima, prikupi njen deljeni sadržaj i zaključi da li su objave pozitivne, negativne ili neutralne. Zatim, pored osnovnih informacija o svakoj objavi pojedinačno, prikazaće i u koju, od tri navedene grupe objava spada.

1. Primena projekta

Sada, kada znamo koji je zadatak aplikacije, mozemo reći da je ovo moderan način za praćenje i proučavanje studenata. Tradicionalni pristup koji se odnosi na usmene razgovore ili pisane formulare, nosi sa sobom velika ograničenja, od kojih je jedno broj studenata koji se mogu ispitati. Upravo ovde vidimo prednost modernog načina koji daje mogućnost da se za kratko vreme obradi velika količina podataka. Bilo koja organizacija, koja zeli da brzo prouči osećanja/raspoloženja/mišljenja određene grupe, može da primeni ovaj projekat.

1. Implementacija

U nastavku se nalaze sve biblioteke koje su korišćene u projektu i koje olakšavaju rad sa Reddit API-jem.

5.1. Requests - biblioteka koja je de facto standard za pravljenje HTTP zahteva u Python-u.

5.2. Pandas - biblioteka koja pruša brze i fleksibilne strukture podataka koje se koriste za rad i manipulaciju podacima.

5.3. Textblob - biblioteka za obradu tekstualnih podataka koja nad njima može da izvodi operacije kao sto su analiza osećanja i klasifikacija.

1. Textblob

Kako je tema projekta praćenje aktivnosti studenata, kakvo je njihovo raspoloženje, odnos prema fakultetu i njihov osećaj, zadržaćemo se malo na biblioteci textblob.

Naveli smo da ona može nad podacima da izvrsava operaciju analiza osećanja, pa hajde da pojasnimo šta je zapravo to.   
Analiza osećanja nam moze pomoci da dešifrujemo raspoloženje i emocije pojedinca ili grupe, i prikupimo pronicljive informacije u vezi sa kontekstom. Analiza osećanje je proces analize podataka i klasifikacije na osnovu potrebe istraživanja.

Textblob je biblioteka za obradu prirodnog jezika(NLP). Za pristupe koji su zasnovani na leksici, osećanja su definisana njenom semantičkom orijentacijom i intenzitetom svake reči u rečenici. To zahteva unapred definisan rečnik koji klasifikuje negativne i pozitivne reči. Kako je recenica predstavljena kao niz reči, nakon dodeljivanja ocena svim rečima, konačni osećaj se izračunava operacijom udruzivanja. Textblob vraća polaritet rečenice. Polaritet se nalazi u opsegu [-1,1]. -1 definiše negativno osećanje, 0 neutralno osećanje, a 1 pozitivno osećanje. Textblob takođe ima i semantičke oznake koje pomažu u detaljnoj analizi, kao što su emotikoni, uzvičnici,..

1. Kod

Neophodni kredencijali za korišćenje Reddit API-ja, koji se generišu na osnovu korisničkog imena i šifre i traju ~2h.



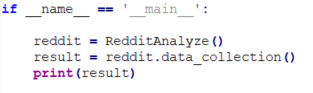
Implementacija klase RedditAuth koja služi za autentifikaciju i pribavljanje access token za pristup Reddit API-ju.



Implementacija klase RedditAnalyze kojom je moguće pribavljanje željenih objava i njihovo analiziranje.



Implementacija metode main unutar koje se koristi objekat klase RedditAnalyze i prikazuje rezultat analize.



1. Instalacije

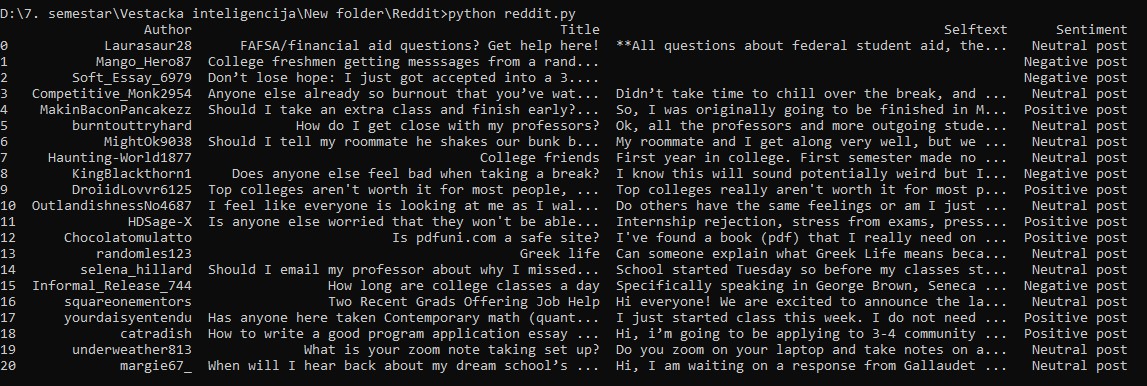
S obzirom da je projekat rađen u python-u, neophodno je na računaru imati instaliran python, kao i korišćene biblioteke (koričćenjem naredbe pip install i navođem imena biblioteka).

1. Pokretanje

Aplikacija se pokreće preko terminala, pozicioniranjem na folder gde se navodi projekat i pozivanjem naredbe ‘python reddit.py’.

1. Rezultat

Na narednoj slici prikazan je rezultat izvršenja aplikacije. Iz navedenih rezultata se vidi da ima 6 pozitivnih, 11 neutralnih i 4 negativne objave.



1. Literatura
2. <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>
3. <https://www.geeksforgeeks.org/python-pandas-dataframe/>
4. <https://www.reddit.com/dev/api>
5. <https://www.reddit.com/prefs/apps>
6. <https://stackoverflow.com/questions/28955541/how-to-get-access-token-reddit-api>