

Projekt 2

Martyna Kaczmarczyk

19 czerwca 2024

Spis treści

1	Wstęp	1
2	Ściąganie postów	1
3	Łączna analiza	2
4	Podsumowanie	5

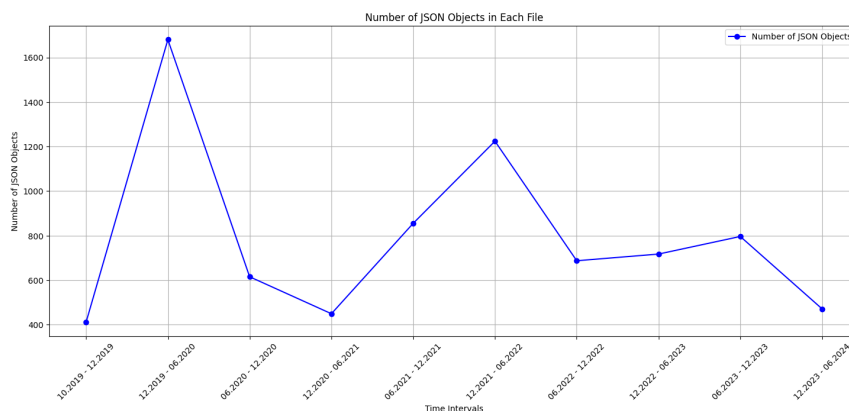
1 Wstęp

Opracownie przedstawia zmieniające się emocje wśród użytkowników portalu Reddit odnośnie serialu 'Witcher', popularność serii oraz zbadane zostało, czy odbiór tego serialu był pozytywny czy negatywny na przestrzeni ostatnich lat. Stworzona przeze mnie baza danych zawiera około 8000 postów, które zostały pobrane z portalu Reddit dzięki bibliotece *praw*. Następnie dane zostały przeanalizowane dzięki bibliotekom takim jak *text2emotion*, *nlTK*, *wordcloud* i *transformers*.

2 Ściąganie postów

Ustawiłam limit postów dla danego przedziału czasowego na 2000, jednak w żadnym okresie programowi nie udało się ściągnąć aż takiej liczby. Program pobierał dane z subredditów: *'netflixwitcher'*, *'witcher'*, *'television'*, *'gaming'*, *'cosplay'*, *'fantasy'*, a następnie w znalezionych postach program poszukiwał słów kluczowych (*'Witcher'*, *'witcher'*, *'Wiedmin'*, *'witcher'*, *'Ciri'*, *'Jaskier'*, *'Yarpen'*, *'Geralt'*, *'Yennefer'*, *'Triss'*, *'Nilfgaard'*, *'Tissaia'*, *'Dzikigon'*, *'TimeofContempt'*). Dopiero następnie pobierał informacje i zapisywał do pliku.

Na wykresie poniżej pokazane jest, ile postów udało się mojemu programowi ściągnąć w podanych okresach czasu (każdy okres wynosi pół roku).



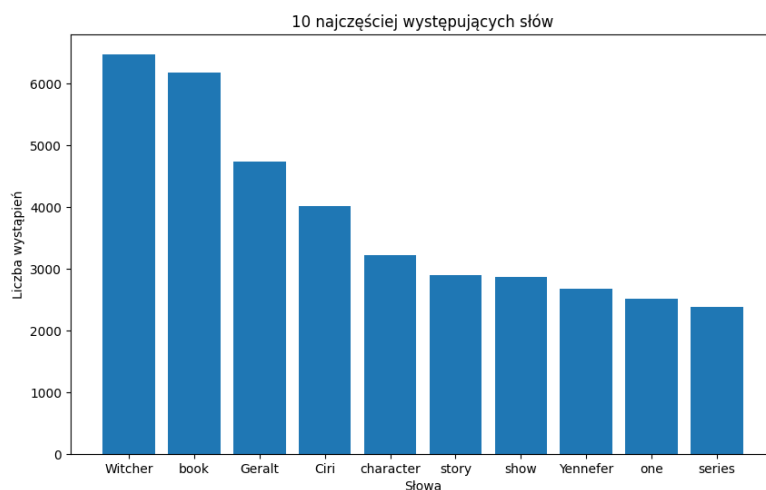
Widać, że użytkownicy wykazywali największe zainteresowanie serią po produkcji pierwszego sezonu serialu na platformie Netflix, która odbyła się 21 grudnia 2019. Potem zainteresowanie spadło, jednak dzięki ogłoszeniom o zbliżającym się drugim sezonie z powrotem zaczęło rosnąć. Jednak nawet po wydaniu kontynuacji serii 17 grudnia 2021 zainteresowanie użytkowników nie było aż tak duże jak po premierze pierwszego sezonu. Następnie, trzeciemu sezonowi towarzyszyła jeszcze mniejsza uwaga od strony użytkowników platformy Reddit. Na chwilę obecną wykres pokazuje, że seria nie skupia wokół siebie dużego zainteresowania.

3 Łączna analiza

Poniższe wyniki to wyniki ogółe dla wszystkich pobranych postów. Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 1079669, a po przeprowadzeniu lematyzacji - 463295.



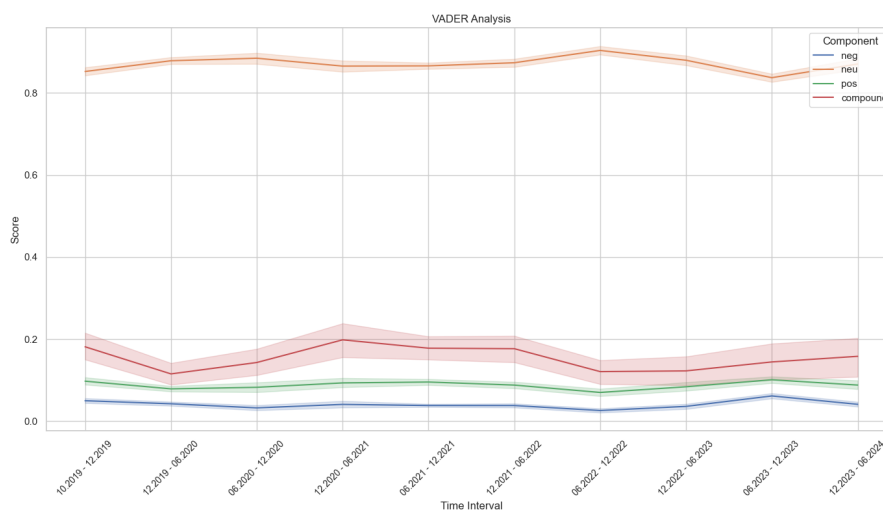
Wykres 10-ciu najczęstszych słów



Nie jest zaskakujące to, że najpopularniejszym słowem jest słowo *Witcher*. Jednak drugim popularnym słowem jest *book*, co może oznaczać, że książki Andrzeja Sapkowskiego były popularnym tematem postów. Oczywiście, w zaprezentowanym top-10 pojawiają się także imiona głównych bohaterów.

Na poniższych wykresach linie to przypuszczany wynik, a jaśniejsze pola dookoła nich to możliwy błąd obliczeń.

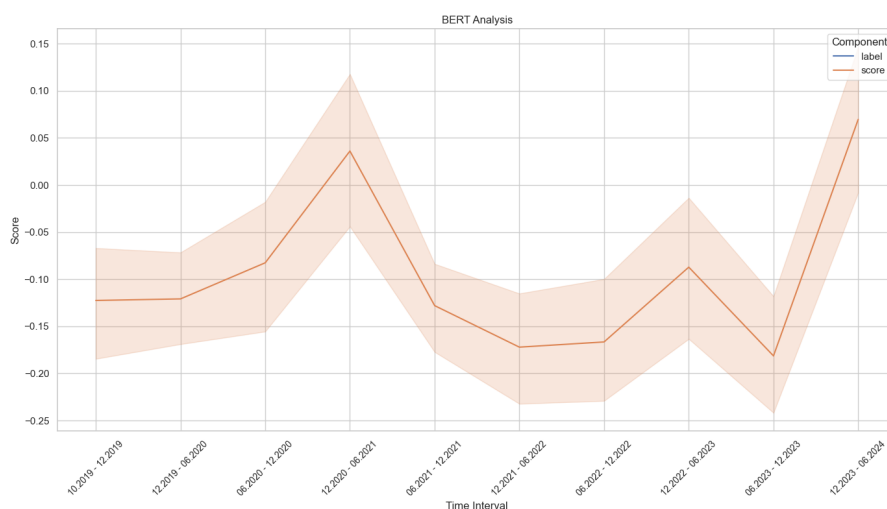
Poniższy wykres przedstawia jak zmienia się średnia dla każdego komponentu VADER'a na przedziałach czasu.



Widać, że algorytm był najbardziej pewny neutralnego wydźwięku w postach. Następnie, najpewniej VADER określił się w stosunku do opinii mie-

szanych, jednak jest to o wiele mniejszy *score* niż ten o opinii neutralnej. Można też zauważyć, że algorytm jest bardziej pewny pozytywnych wydźwięków postów, niż negatywnych. Przy czym wszystkie krzywe są w miarę stałe.

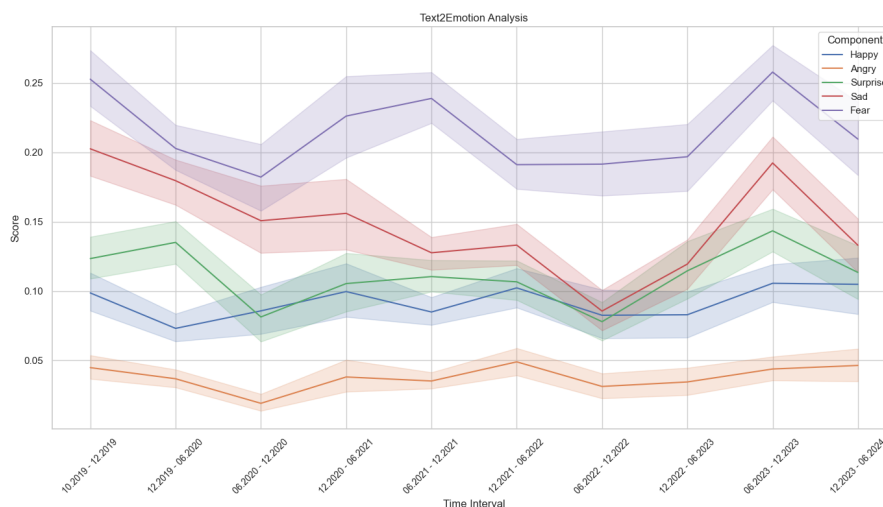
Poniższy wykres przedstawia jak zmienia się średnia dla algorytmu BERT na przedziałach czasu, czyli została obliczona średnia ze *score* dla każdego okresu dla opinii negatywnych (wartość -1) i pozytywnych (wartość 1)



Na wykresie można zobaczyć, że na początku (w październiku 2019 roku) posty według tego algorytmu mają wydźwięk trochę negatywny, jednak krzywa ma tendencję rosnącą. Uda jej się przekroczyć próg 0.0 sięgając lekko ponad, zanim zacznie maleć, ponownie wracając do lekko negatywnych wydźwięków. Pozostaje po tej stronie aż do grudnia 2023 roku. Po tym średnia *score* gwałtownie rośnie. Na dzień dzisiejszy (czerwiec 2024) seria *Wiedmina* wydaje się być pozytywnie oceniana przez użytkowników Reddita.

Co ciekawe algorytmy BERT i VADER pokazują dwie tendencje rosnące krzywych w tych samych obszarach czasu. Pierwszy okres to druga połowa 2020 roku. Jest to czas, kiedy programowi udało się ściągnąć stosunkowo małą liczbę postów, co odzwierciedla popularność serii. Drugi okres to czerwiec 2024 roku (dzień dzisiejszy), który wykazuje się podobnymi cechami, co pierwszy okres. Ten nagły wzrost pozytywnych opinii może wynikać z tego, że w tym okresie przy serii zostali najwierniejsi fani *Wiedmina*, a wszelcy potencjalni krytycy zajęli się wystawianiem negatywnych opinii innym seriom.

Poniższy wykres pokazuje wyliczone średnie każdej z emocji dla podanych okresów.



Można zauważyć, że krzywe rosną i maleją, jednak wciąż zachowują pewną ustaloną kolejność. Strachowi zawsze towarzyszy najwyższy współczynnik. Współczynnik smutku jest zawsze troszkę niższy, jednak wciąż wyższy od wszelkich innych emocji w większości przypadków. Tylko pierwszej połowie 2022 roku ma wartość podobną do tej należącej do krzywej zaskoczenia i szczęścia. Z kolei krzywe zaskoczenia i szczęścia prawie zawsze się na siebie nakładają lub przeplatają się. Z wykresu można też wyczytać, że prawdopodobieństwo pojawienia się w tekstach złości jest najniższe we wszystkich okresach, co ogólnie może przyjąć za wynik pozytywny.

4 Podsumowanie

Można zauważyć, że popularność serii *Wiedmin* jest zależna od danego okresu oraz wydarzeń mających w nim miejsce. Okazuje się, że seria była najbardziej popularna, kiedy na platformie *Netflix* odbyła się premiera pierwszego sezonu. Nigdy to już się nie potoczyło, jednak zostali jej lojalni najwięksi fani. Dzięki temu seria w najbardziej spokojnych okresach była głównie chwalona, nie krytykowana.