

Projekt 2

Martyna Kaczmarczyk

19 czerwca 2024

Spis treści

1 Wstęp	2
2 Ściąganie postów	3
3 Łączna analiza	4
4 Szczegółowa analiza każdego okresu	7
4.1 10.2019 - 12.2019	7
4.1.1 Bag of words i Word cloud	7
4.1.2 Analiza odbioru w określonym czasie	8
4.1.3 Analiza wyników przetworzonych przez VADER'a	9
4.1.4 Analiza wyników przetworzonych przez BERT'a	9
4.1.5 Dystrybucja przypisanych emocji	10
4.2 12.2019 - 06.2020	11
4.2.1 Bag of words i Word cloud	11
4.2.2 Analiza odbioru w określonym czasie	12
4.2.3 Analiza wyników przetworzonych przez VADER'a	12
4.2.4 Analiza wyników przetworzonych przez BERT'a	13
4.2.5 Dystrybucja przypisanych emocji	14
4.3 06.2020 - 12.2020	14
4.3.1 Bag of words i Word cloud	15
4.3.2 Analiza odbioru w określonym czasie	15
4.3.3 Analiza wyników przetworzonych przez VADER'a	16
4.3.4 Analiza wyników przetworzonych przez BERT'a	16
4.3.5 Dystrybucja przypisanych emocji	17
4.4 12.2020 - 06.2021	17
4.4.1 Bag of words i Word cloud	18
4.4.2 Analiza odbioru w określonym czasie	19
4.4.3 Analiza wyników przetworzonych przez VADER'a	19
4.4.4 Analiza wyników przetworzonych przez BERT'a	20
4.4.5 Dystrybucja przypisanych emocji	21
4.5 06.2021 - 12.2021	21

4.5.1	Bag of words i Word cloud	22
4.5.2	Analiza odbioru w określonym czasie	22
4.5.3	Analiza wyników przetworzonych przez VADER'a	23
4.5.4	Analiza wyników przetworzonych przez BERT'a	24
4.5.5	Dystrybucja przypisanych emocji	24
4.6	12.2021 - 06.2022	25
4.6.1	Bag of words i Word cloud	25
4.6.2	Analiza odbioru w określonym czasie	26
4.6.3	Analiza wyników przetworzonych przez VADER'a	27
4.6.4	Analiza wyników przetworzonych przez BERT'a	27
4.6.5	Dystrybucja przypisanych emocji	28
4.7	06.2022 - 12.2022	28
4.7.1	Bag of words i Word cloud	29
4.7.2	Analiza odbioru w określonym czasie	30
4.7.3	Analiza wyników przetworzonych przez VADER'a	30
4.7.4	Analiza wyników przetworzonych przez BERT'a	31
4.7.5	Dystrybucja przypisanych emocji	31
4.8	12.2022 - 06.2023	32
4.8.1	Bag of words i Word cloud	32
4.8.2	Analiza odbioru w określonym czasie	33
4.8.3	Analiza wyników przetworzonych przez VADER'a	34
4.8.4	Analiza wyników przetworzonych przez BERT'a	34
4.8.5	Dystrybucja przypisanych emocji	35
4.9	06.2023 - 12.2023	35
4.9.1	Bag of words i Word cloud	35
4.9.2	Analiza odbioru w określonym czasie	36
4.9.3	Analiza wyników przetworzonych przez VADER'a	37
4.9.4	Analiza wyników przetworzonych przez BERT'a	37
4.9.5	Dystrybucja przypisanych emocji	38
4.10	12.2023 - 06.2024	38
4.10.1	Bag of words i Word cloud	39
4.10.2	Analiza odbioru w określonym czasie	39
4.10.3	Analiza wyników przetworzonych przez VADER'a	40
4.10.4	Analiza wyników przetworzonych przez BERT'a	40
4.10.5	Dystrybucja przypisanych emocji	41
5	Podsumowanie	41

1 Wstęp

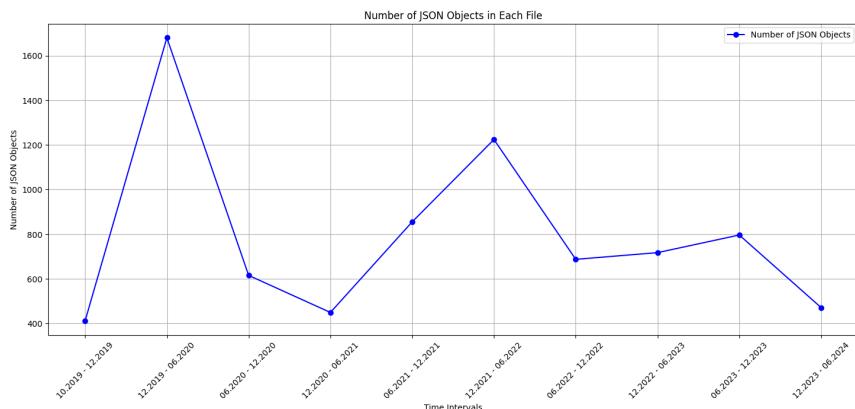
Opracowanie przedstawia zmieniające się emocje wśród użytkowników portalu Reddit odnośnie serialu 'Witcher', popularność serii oraz zbadane zostało,

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 Ściaganie postów

Ustawiłem limit postów dla danego przedziału czasowego na 2000, jednak w żadnym okresie programowi nie udało sięściągnąć aż takiej liczby. Program pobierał dane z subredditów: '*netflixwitcher*', '*witcher*', '*television*', '*gaming*', '*cosplay*', '*fantasy*', a następnie w znalezionejnych 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.

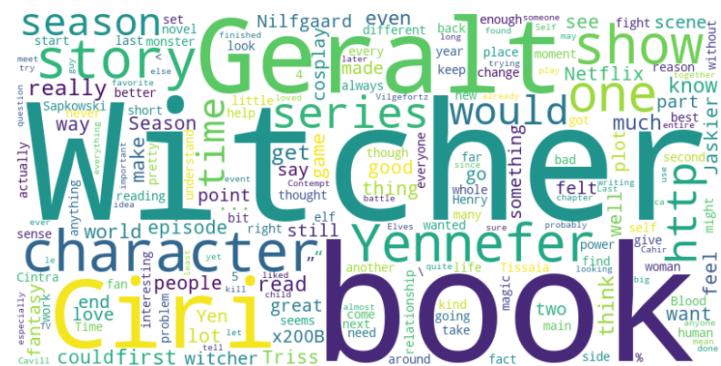
Na wykresie poniżej pokazane jest, ile postów udało się mojemu programowiściągnąć w podanych okresach czasu (każdy okres wynosi pół roku).



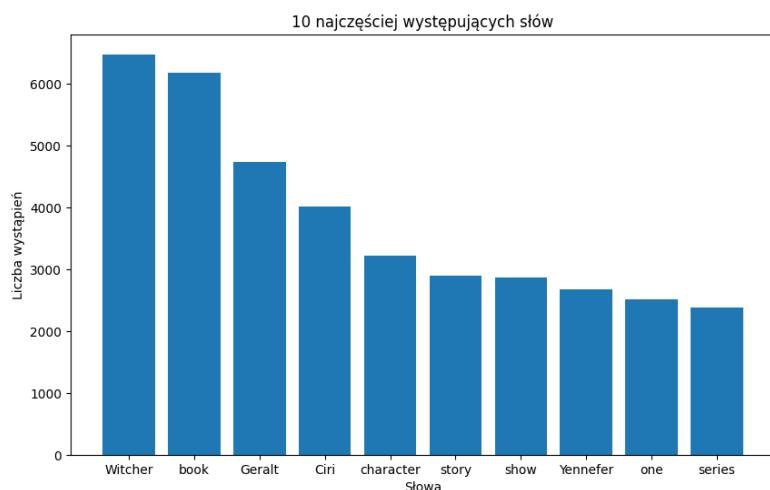
Widać, że użytkownicy wykazywali największe zainteresowanie serią po produkcji pierwszego sezonu serialu na platformie Netflix, która odbyła się 21 grudnia 2019. Potem zainteresowanie spadło, jednak dzięki ogłoszeniom o zbliżającym się drugim sezonie z powrotem zaczęło rosnąć. Jednak nawet po wydaniu 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 ogółem dla wszystkich pobranych postów. Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 1079669, a po przeprowadzeniu lematyzacji - 463295.



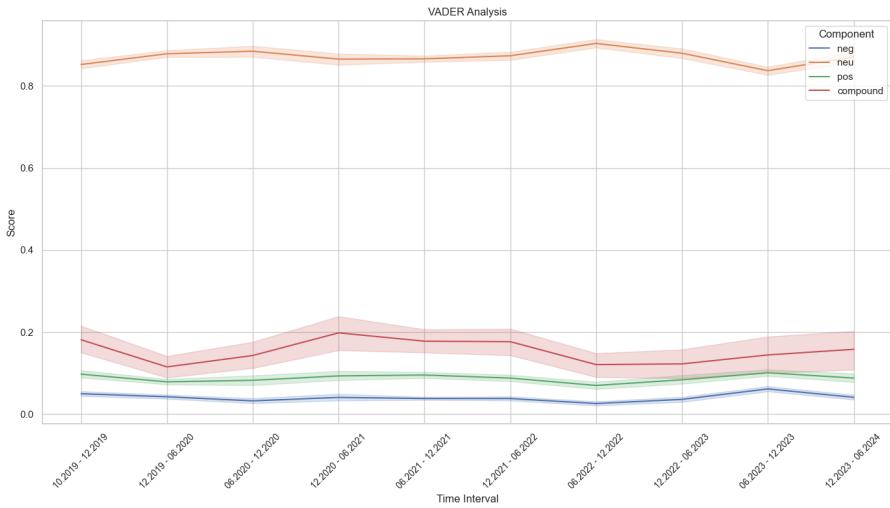
Wykres 10-ciu najczęstszych słów



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.

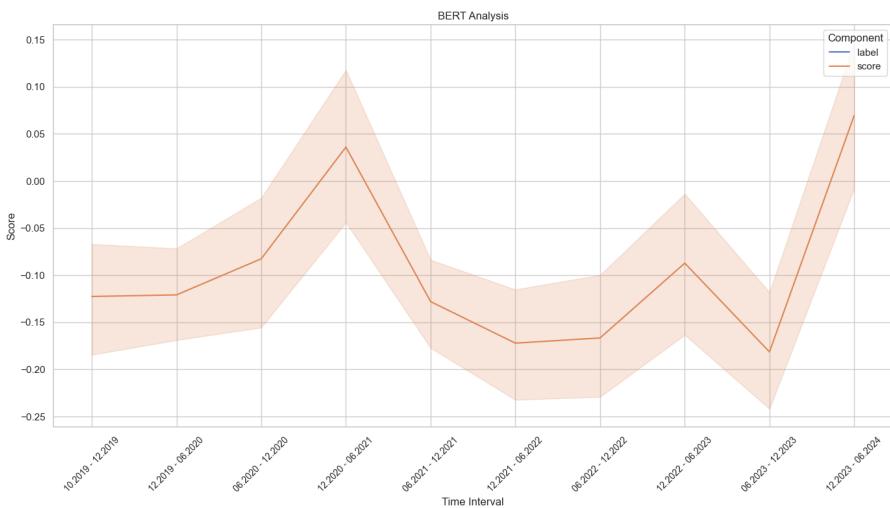
Na poniższych wykresach linie to przypuszczany wynik, a jaśniejsze pola dookoła nich to możliwy błąd obliczeń.

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



Widać, że algorytm był najbardziej pewny neutralnego wydawniku w postach. Następnie, najpewniej VADER określił się w stosunku do opinii mieszanych, jednak jest to o wiele mniejszy *score* niż ten o opinii neutralnej. Można też zauważyc, że algorytm jest bardziej pewny pozytywnych wydawników postów, niż negatywnych. Przy czym wszystkie krzywe są w miarę stałe.

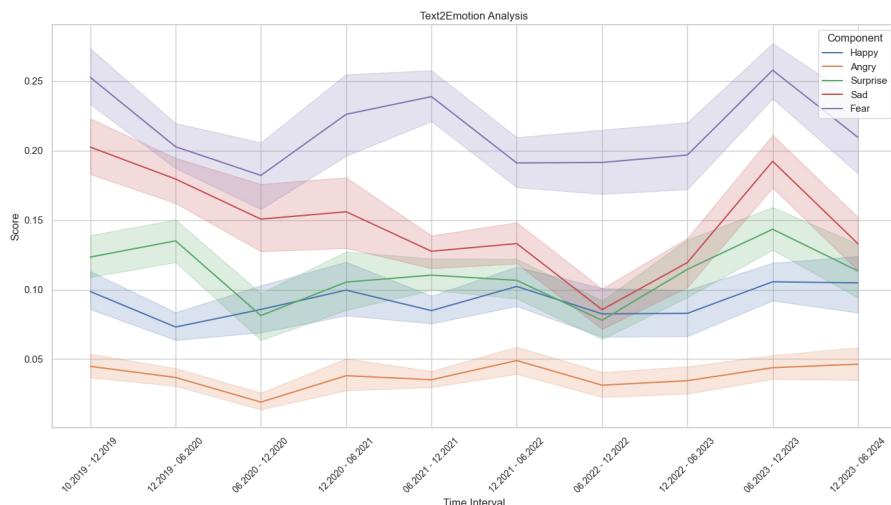
Poniższy wykres przedstawia jak zmienia się średnia dla algorytmu BERT na przedziałach czasu, czyli została obliczona średnia ze *score* dla każdego okresu dla opinii negatywnych (wartość -1) i pozytywnych (wartość 1)



Na wykresie można zobaczyć, że na początku (w październiku 2019 roku) posty według tego algorytmu mają wydźwięk trochę negatywny, jednak krzywa ma tendencję rosnącą. Udaje 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 Redditu.

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 (dzien 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 zasze się na siebie nakładają lub przeplatają się. Z wykresu można też wyczytać, że prawdopodobieństwo pojawienia się w tekstu złości jest najniższe we wszystkich

okresach, co ogólnie można przyjąć za wynik pozytywny.

4 Szczegółowa analiza każdego okresu

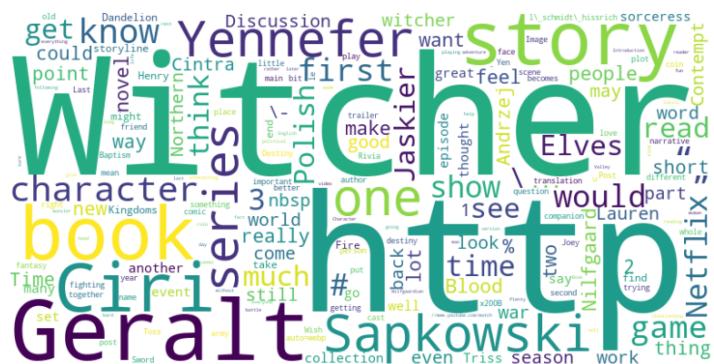
A teraz przyjrzymy się bliżej jak seria sobie radziła w poszczególnych okresach.

4.1 10.2019 - 12.2019

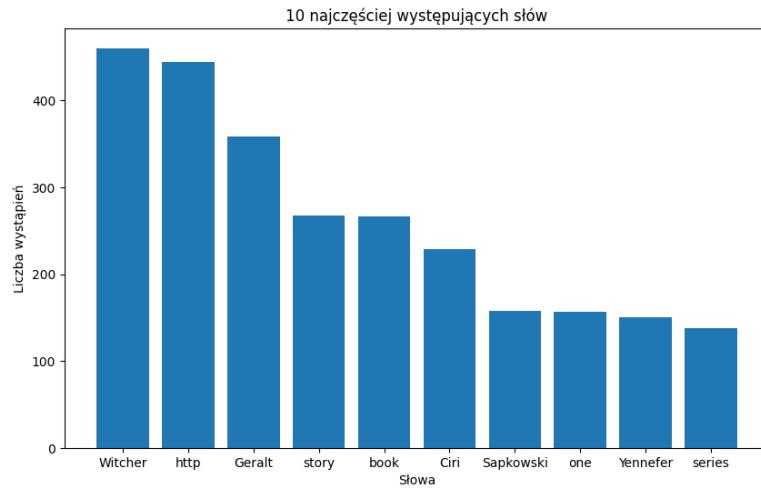
Pierwszym omawianym okresem jest czas przed wyjściem pierwszego sezonu serialu na platformie Netflix.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 74855, a po przeprowadzeniu lematyzacji - 33721.

4.1.1 Bag of words i Word cloud



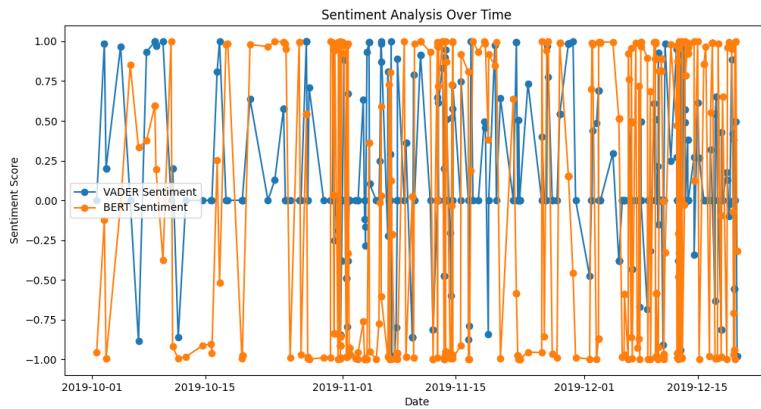
Wykres 10-ciu najczęstszych słów



Jak widać na wykresie najczęściej pojawiającym się słowem jest *Witcher*, co raczej nie jest zaskoczeniem. Na drugim miejscu jest *http*, co może znaczyć, że autorzy postów często odwoywali się do linków. I dopiero na trzecim miejscu można znaleźć imię głównego bohatera.

4.1.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.

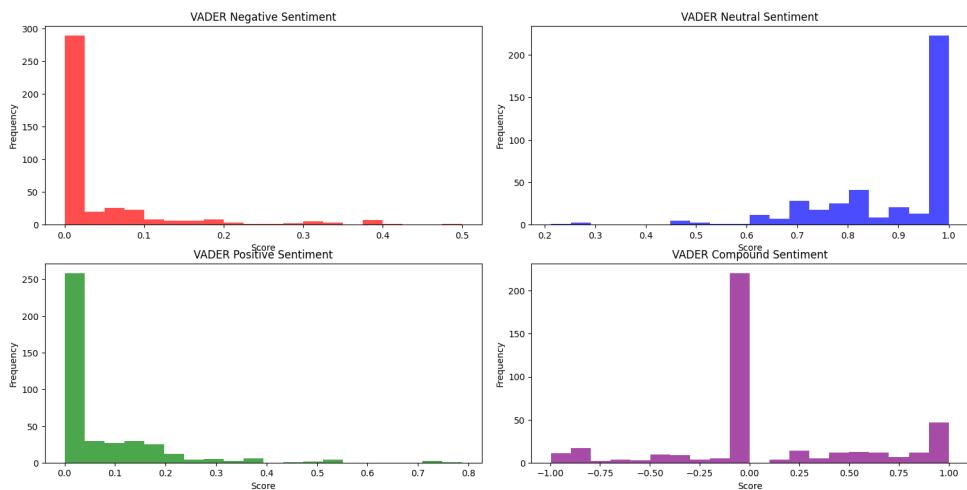


Jak widać seria tutaj cieszyła się największą uwagą na początku listopada 2019 roku oraz w grudniu. Można też zauważyc, że zawsze można znaleźć ludzi pozytywnie i negatywnie nastawionych. Należy też zwrócić uwagę na to, że BERT sentiment w większości przypadków jednoznacznie określa tekst

za pozytywny lub negatywny, gdzie VADER sentiment często rozpoznaje wiele postów za neutralne.

4.1.3 Analiza wyników przetworzonych przez VADER'a

Poniższe histogramy pokazują, jakie wartości przyjęły podstawowe komponenty.

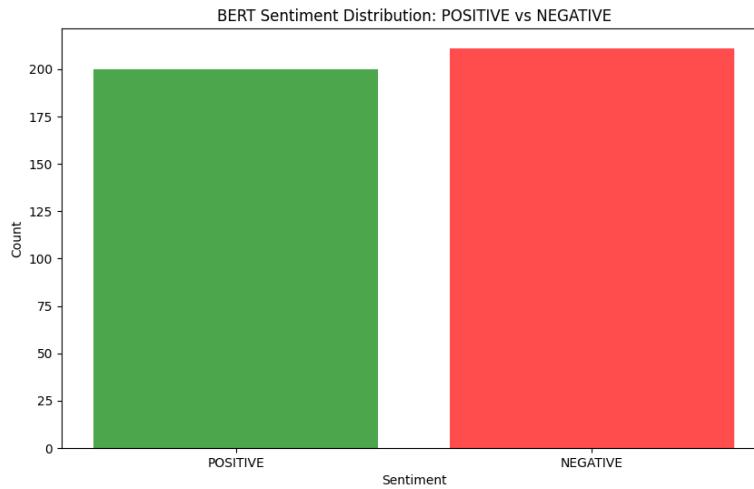


VADER przypisuje każdemu komponentowi pewien procent prawdopodobieństwa. W pierwszym histogramie, dotyczącym postów ocenionych negatywnie, najwyższy słupek osiąga prawie 300 przy niskim procencie prawdopodobieństwa, niewiele większym od 0. Histogram pozytywnego komponentu wykazuje podobny trend, z najwyższym słupkiem osiągającym wartość około 250 również przy niskim procencie prawdopodobieństwa. Jednakże, ten histogram wskazuje również, że istnieją pozytywnie ocenione posty, których procent prawdopodobieństwa jest większy, sięgając do 0.2.

Idąc dalej, histogram postów ocenionych jako neutralne jest bardzo wyraźny. Większość tekstów w tej kategorii osiąga wartość bliską 1, co oznacza, że są one oceniane jako zdecydowanie neutralne. Z kolei ostatni histogram, dotyczący ocen mieszanych, wskazuje, że znaczna liczba postów ma delikatny negatywny wydźwięk, ponieważ większość wartości *compound* znajduje się nieco poniżej 0. Drugim najwyższym słupkiem na wykresie, choć znacznie mniejszym, jest słupek odpowiadający wartości 1, co oznacza, że VADER ocenił te posty jako zdecydowanie pozytywne.

4.1.4 Analiza wyników przetworzonych przez BERT'a

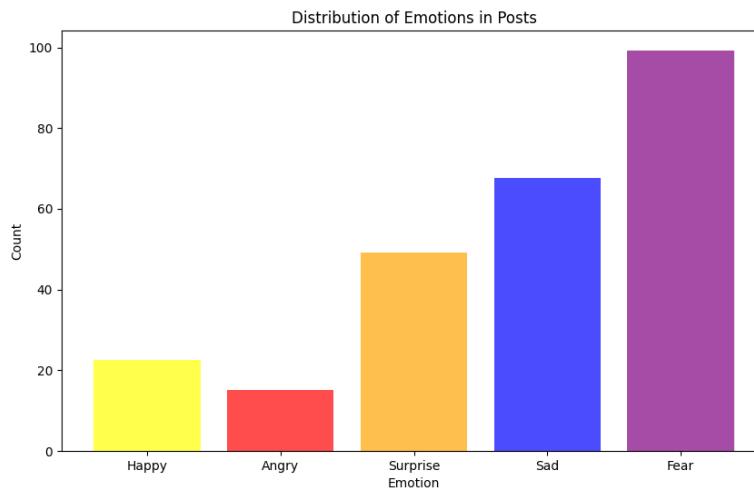
Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



Liczba postów określonych za negatywne jest trochę większa. Może to wynikać z obaw fanów książek 'Wiedźmin' przed powstaniem kolejnej ekranizacji powieści.

4.1.5 Dystrybucja przypisanych emocji

Poniższy histogram pokazuje, ile emocji zostało określonych we wszystkich postach. Można zauważyc, że najwyższym słupkiem jest *fear*, czyli strach. Może to wynikać, jak już wcześniej zostało wspomniane, z obawy, jak platformie Netflix uda się zekranizować ukochaną serię społeczności.



4.2 12.2019 - 06.2020

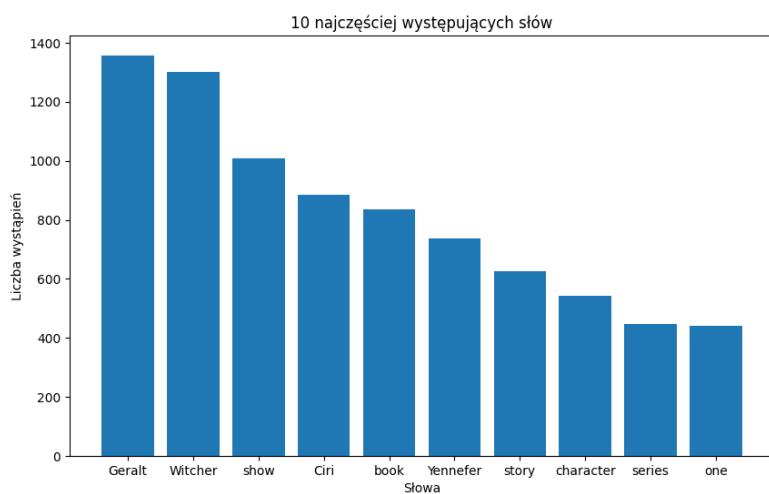
Drugim omawianym okresem jest połowa roku zaraz po wyjściu pierwszego sezonu serialu na platformie Netflix.

Najpierw dane zostały przetworzone. Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 219173, a po przeprowadzeniu lematyzacji - 94692.

4.2.1 Bag of words i Word cloud



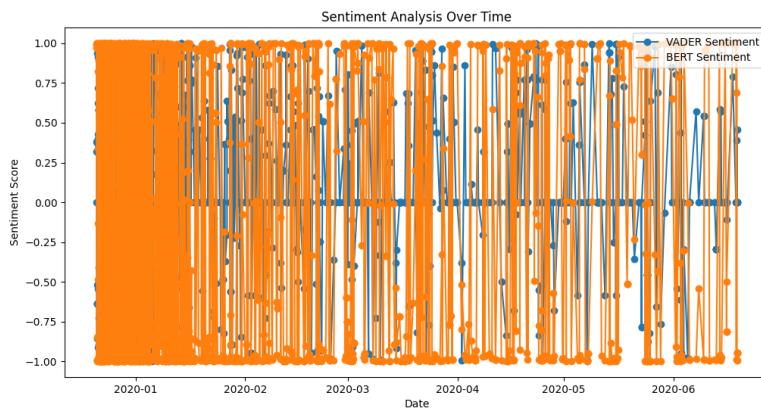
Wykres 10-ciu najczęstszych słów



Jak widać na wykresie najczęściej pojawiającym się słowem jest imię głównego bohatera serii. Na drugim miejscu, choć różnica w ilości jest niewielka, jest tytuł - *Witcher*. Na trzecim miejscu możemy znaleźć wyraz *show*, co pokazuje, że w tym okresie wiele ludzi odwołuje się do premiery nowego serialu.

4.2.2 Analiza odbioru w określonym czasie

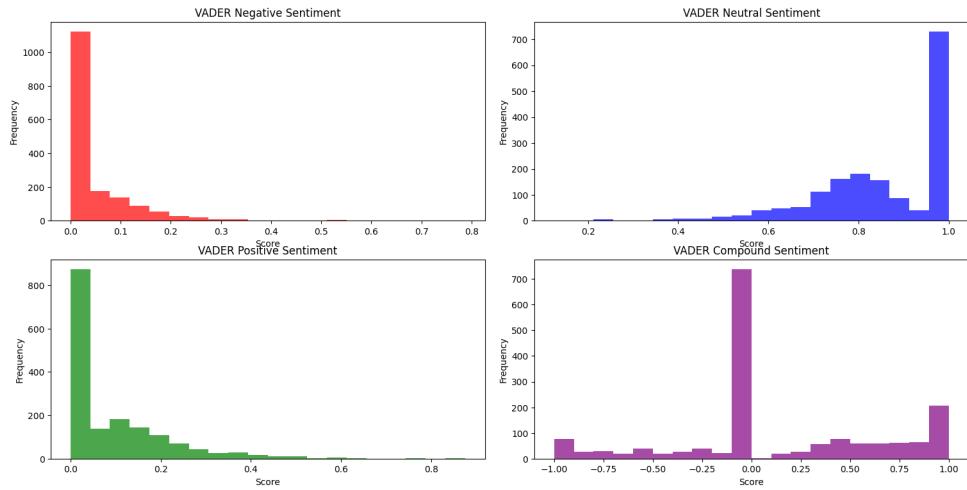
Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



W omawianym okresie temat *Wiedmina* był najczęstszy. Jednak, gdy się bliżej przyjżymy, najwięcej postów pochodzi z końca grudnia 2019 roku i początku stycznia 2020 roku, zaraz po premierze pierwszego sezonu. Widać też, że opinie wydają się być mieszane, przy czym znowu wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny.

4.2.3 Analiza wyników przetworzonych przez VADER'a

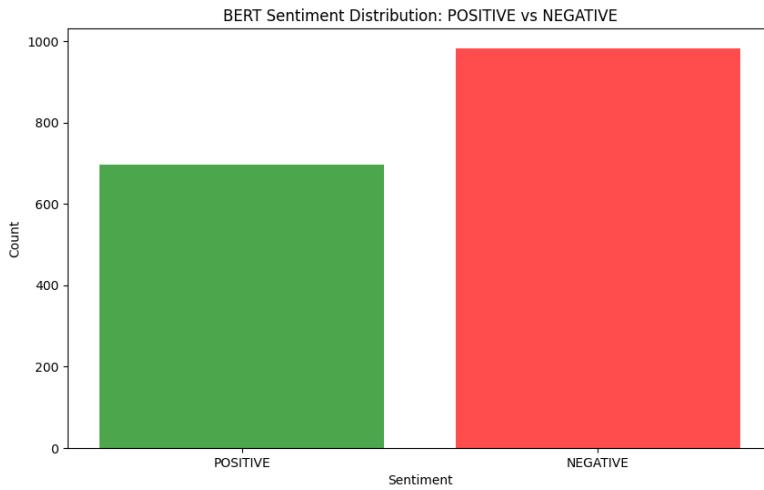
Poniższe histogramy pokazują, jakie wartości przyjęły podstawowe komponenty.



Rozłożenie VADER'a jest podobne, co w poprzednim przypadku. Różnicą jest większa liczba danych.

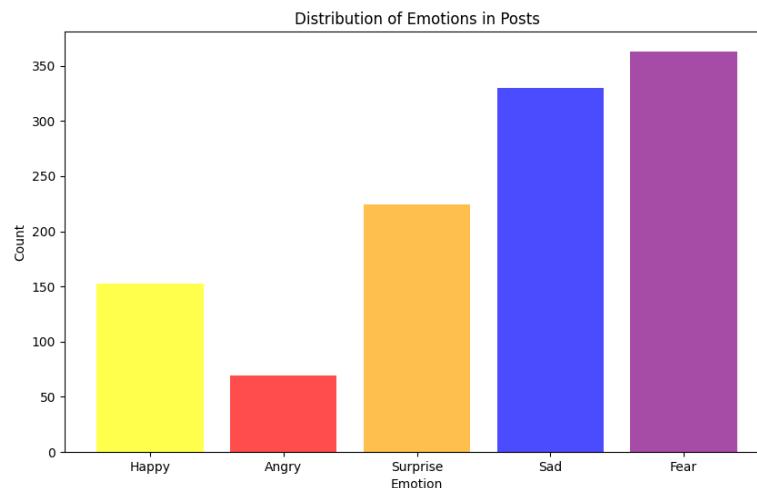
4.2.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



Widać, że BERT przypisał o wiele więcej negatywnych etykieta niż pozytywnych. Etykiet pozytywnych jest około 700, a negatywnych - 1000

4.2.5 Dystrybucja przypisanych emocji



Podobnie jak wcześniej można zauważyc, że najwyższym słupkiem jest strach. Widać też, że tekstem towarzyszy smutek oraz zaskoczenie. Najmniej algorytm przypisał etykietę złości, co może znaczyć, że odbior nie był taki zły.

4.3 06.2020 - 12.2020

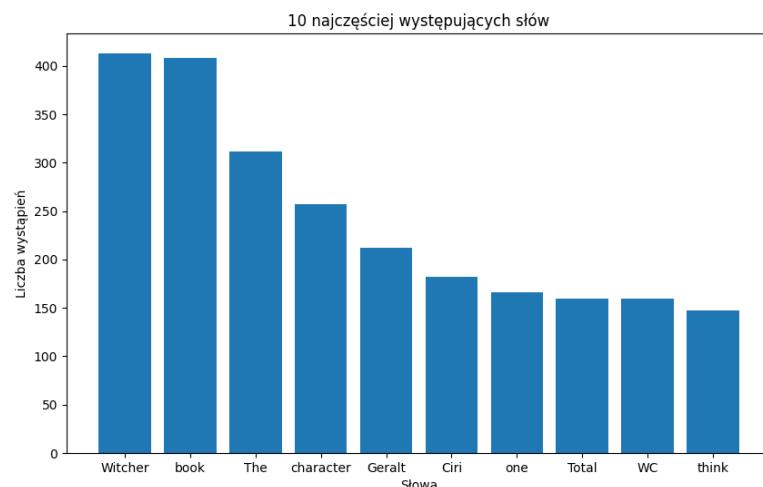
W tym okresie emocje zaczynają opadać i ludzie powoli zapominają o Wiedźminie.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 74855, a po przeprowadzeniu lematyzacji - 33721.

4.3.1 Bag of words i Word cloud



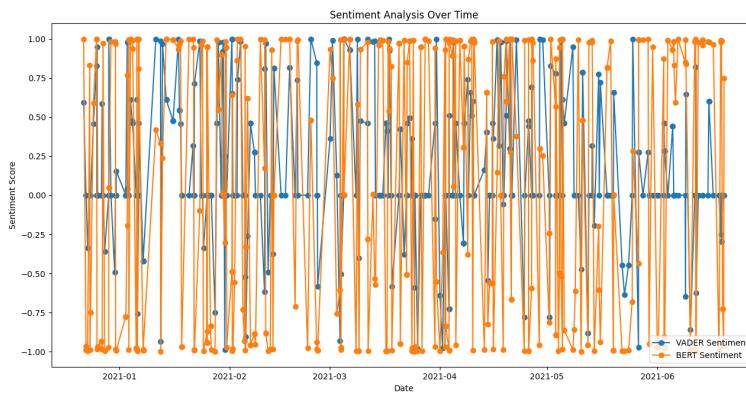
Wykres 10-ciu najczęstszych słów



Jak widać na wykresie najczęściej pojawiającym się słowem jest *Witcher*. Na drugim miejscu jest *book*, co znaczy, że fani zaczynają interesować się książkami Andrzeja Sapkowskiego.

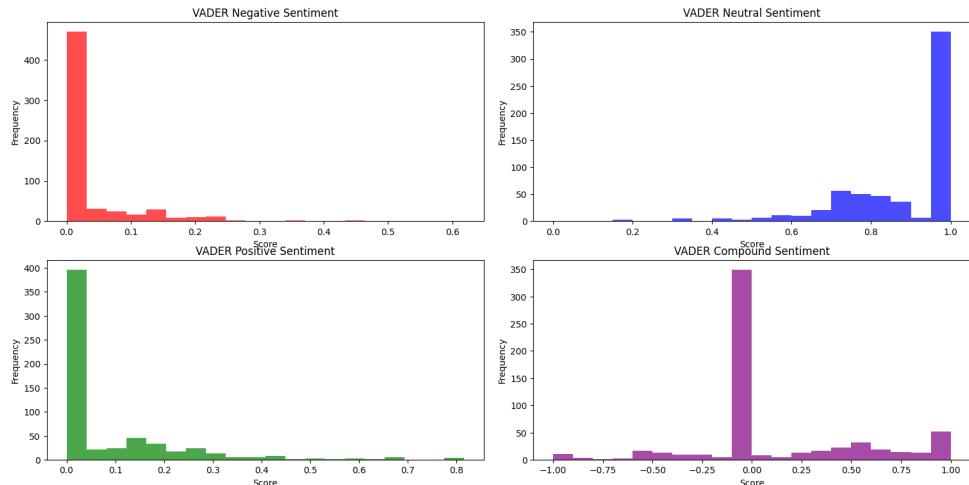
4.3.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagięsczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Można zauważyć, że rozkład postów w tej połowie roku jest w miarę równomierny. Widać też, że opinie wydają się być znowu mieszane i, tak samo jak wcześniej, wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny.

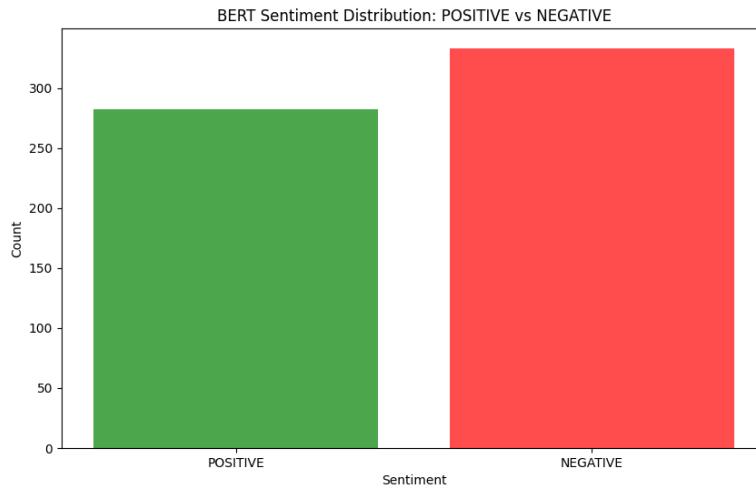
4.3.3 Analiza wyników przetworzonych przez VADER'a



Rozłożenie VADER'a jest podobne, co w poprzednich przypadkach. Komponenty neutralne są najbardziej zdecydowane. Z kolei komponenty negatywne są najbardziej skupione prze wartości niewiele większej od 0.

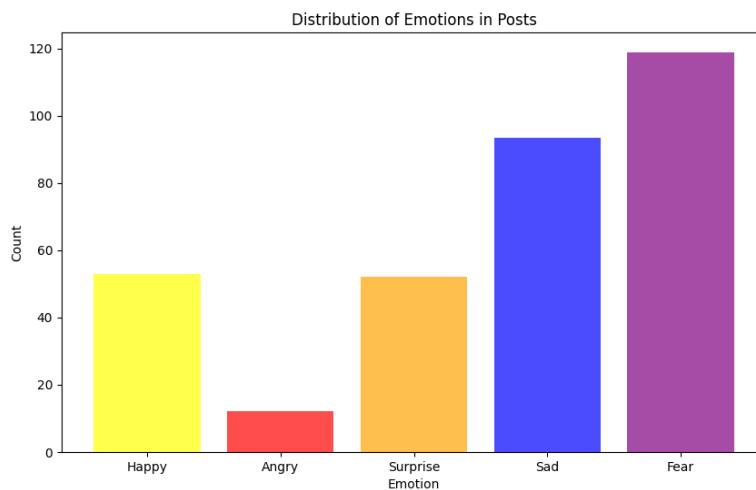
4.3.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



Widać, że BERT ponownie przypisał więcej negatywnych etykiet niż pozytywnych. Etykiet pozytywnych jest około 275, a negatywnych - powyżej 350.

4.3.5 Dystrybucja przypisanych emocji



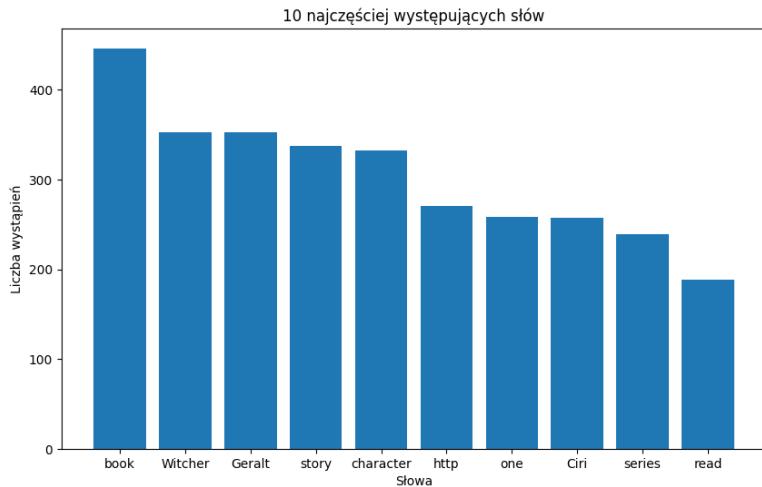
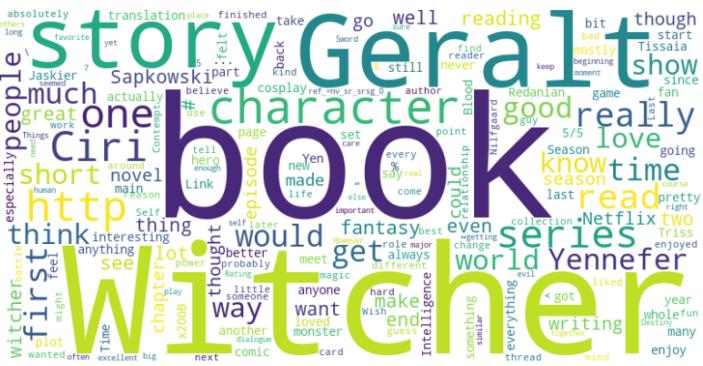
Podobnie jak wcześniej można zauważyć, że najwyższym słupkiem jest strach. Widać też, że tekstem towarzyszy smutek. Poziom zaskoczenia jest podobny do poziomu szczęścia. Najrzadziej znadywaną emocją jest złość.

4.4 12.2020 - 06.2021

Dalej liczba zainteresowanych serią maleje.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 77114, a po przeprowadzeniu lematyzacji - 33415.

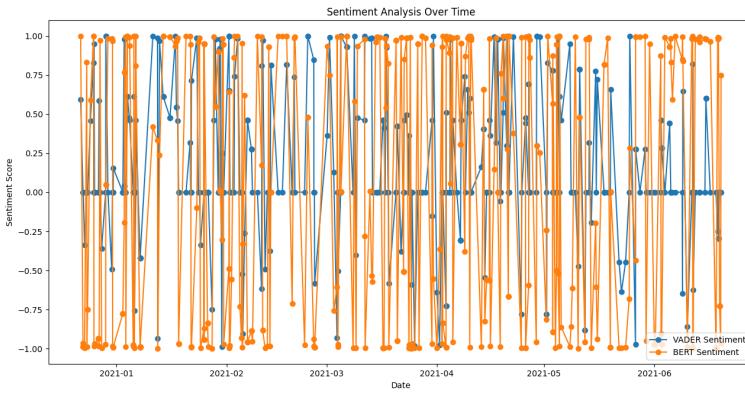
4.4.1 Bag of words i Word cloud



W tym półroczu najczęściej występującym słowem jest *book*. Słowo to prześcignęło nawet *Witcher* i *Geralt*, a co ciekawe w tej rozpisie najczęściej występujących słów nie ma nic o serialu. To może znaczyć, że użytkownicy zajęli się w tym półroczu książką, zapominając o ekranizacji platformy Netflix. Wykres 10-ciu najczęstszych słów

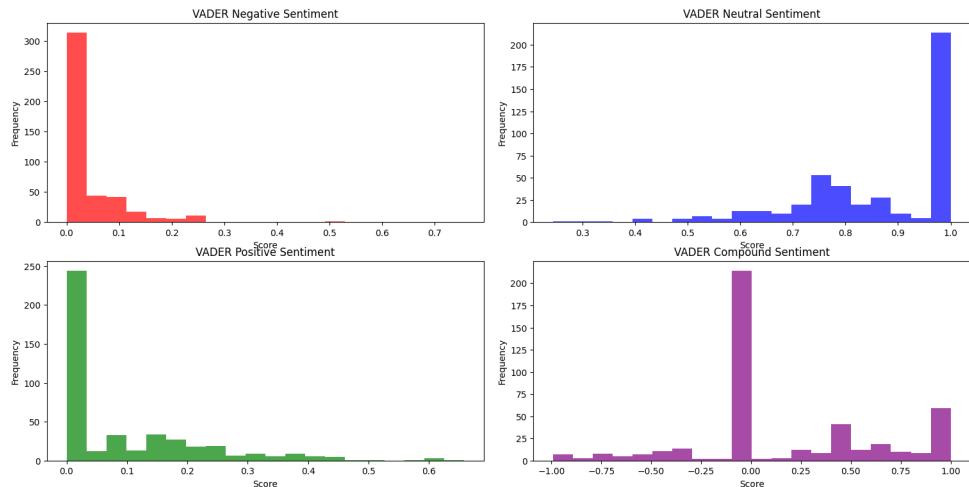
4.4.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zageszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Można zauważyć, że rozkład postów w tej połowie roku jest w miarę równomierny. Widać też, że opinie wydają się być znowu miesiane i, tak samo jak wcześniej, wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny.

4.4.3 Analiza wyników przetworzonych przez VADER'a

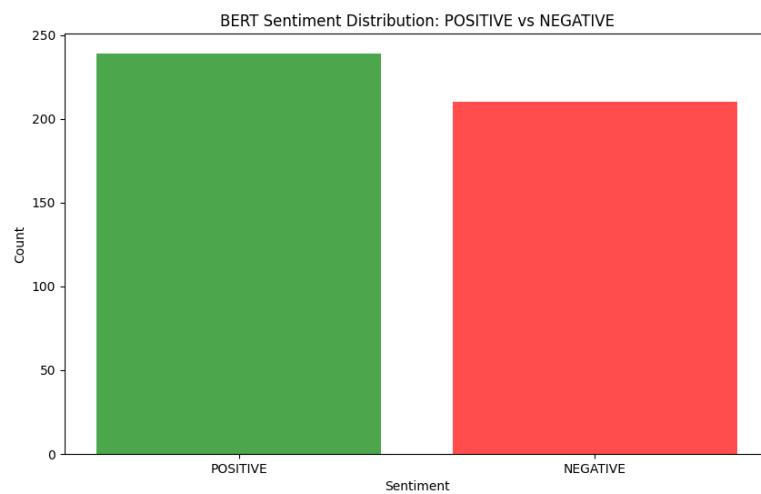


Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach, jednak też widać, że pojawiły się przypadki w pozytywnym histogramie o wyższych wartościach prawdopodobieństwa, które osiągają aż 0.4. Komponenty neutralne są dalej najbardziej zdecydowane, jednak też wzrosły słupki mniej

pewnych przypadków o argumentach 0.7 lub 0.8. Z kolei komponenty negatywne są wciąż najbardziej skupione przy wartości niewiele większej od 0.

4.4.4 Analiza wyników przetworzonych przez BERT'a

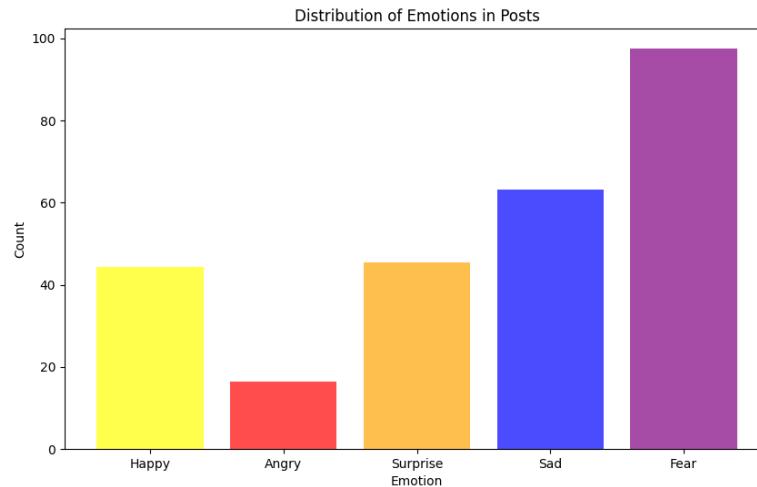
Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie mamy zmianę, bo BERT ocenił więcej tekstów pozytywnie (około 250). Postów ocenionych negatywnie jest około 200.

Chyba z tej części można wnioskować, że fani wolą książkę od serialu.

4.4.5 Dystrybucja przypisanych emocji



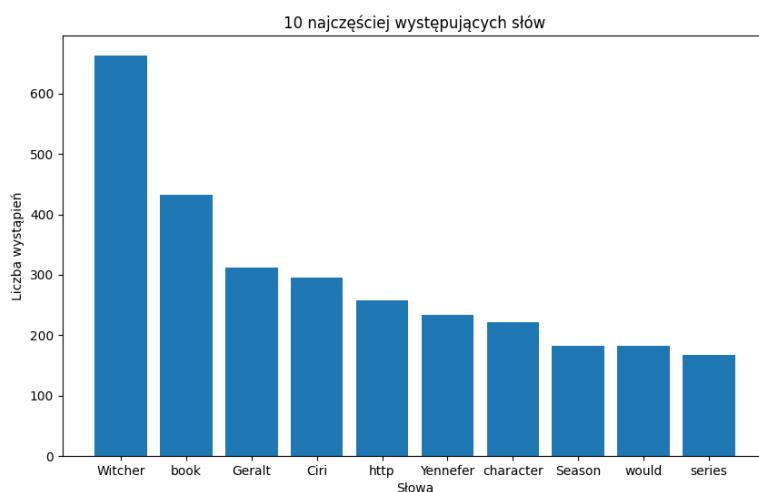
Dystrybucja emocji jest bardzo podobna do poprzedniej. Większość postów ma strachliwy wydźwięk, co może też być spowodowane tym, że seria *Wiedźmin* ma elementy horroru.

4.5 06.2021 - 12.2021

To jest okres przed premierą drugiego sezonu serialu Netflix'a, w którym też liczba postów wzrosła.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 78586, a po przeprowadzeniu lematyzacji - 33821.

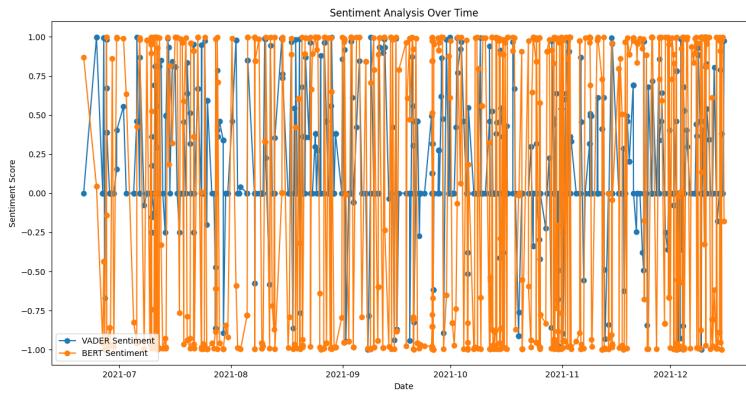
4.5.1 Bag of words i Word cloud



W tym półroczu na pierwsze miejsce wróciło słowo *Witcher*, spyając wyraz *book* na drugie miejsce. W tym zestawieniu można zobaczyć też słowo *Season*, co może znaczyć, że ludzie dzielili się swoimi opiniami odnośnie zbliżającej się premiery.

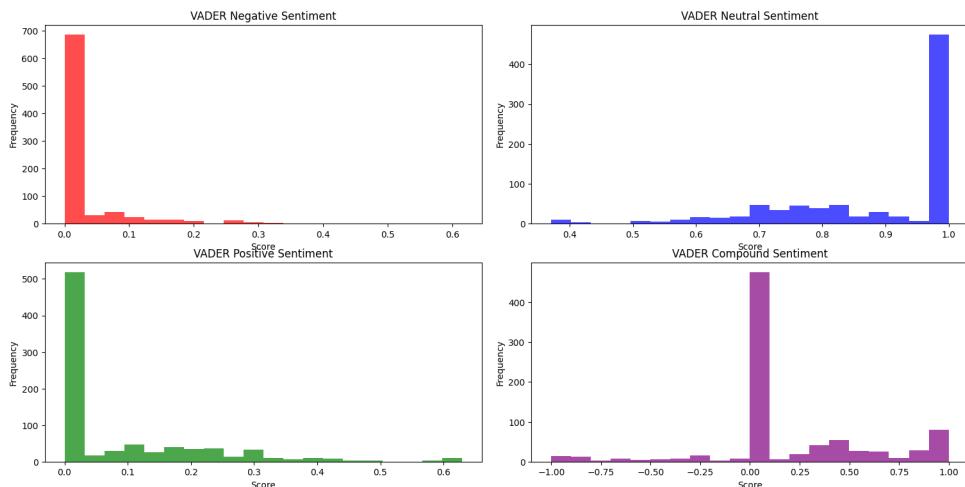
4.5.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Dalej rozkład postów w tej połowie roku jest w miarę równomierny, jednak na początku grudnia tuż przed premierą wykres zaczyna się zagięszczać, czyli liczba postów w tym okresie zaczyna rosnąć. Widać też, że opinie wydają się być dalej mieszane i, tak samo jak wcześniej, wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny. Wydaje mi się też, że VADER w tym okresie ocenia więcej postów pozytywnie, niż negatywnie.

4.5.3 Analiza wyników przetworzonych przez VADER'a

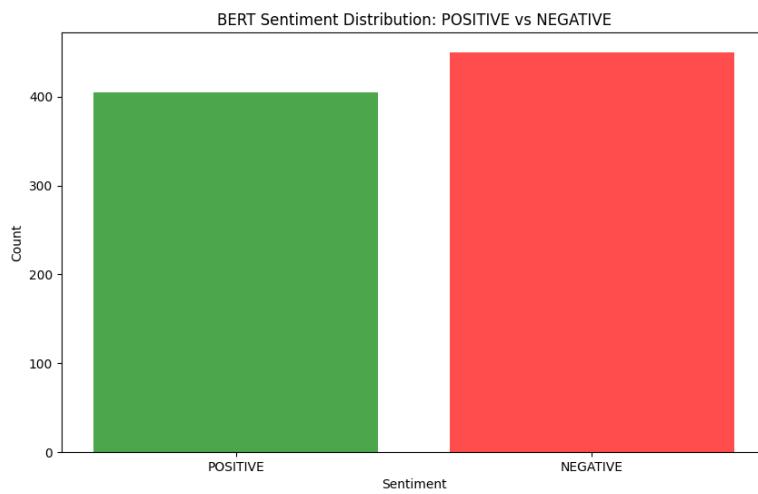


Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach, jednak też widać, że pojawiły się przypadki w pozytywnym histogramie o wyższych wartościach prawdopodobieństwa, które osiągają aż 0.3. To wynik trochę mniejszy niż w ostatnim półroczu. Komponenty neutralne są dalej najbardziej zdecydowane, a słupki mniej pewnych przypadków o argumentach

0.7 lub 0.8 zmalały. Z kolei komponenty negatywne są wciąż najbardziej skupione przy wartościach niewiele większych od 0.

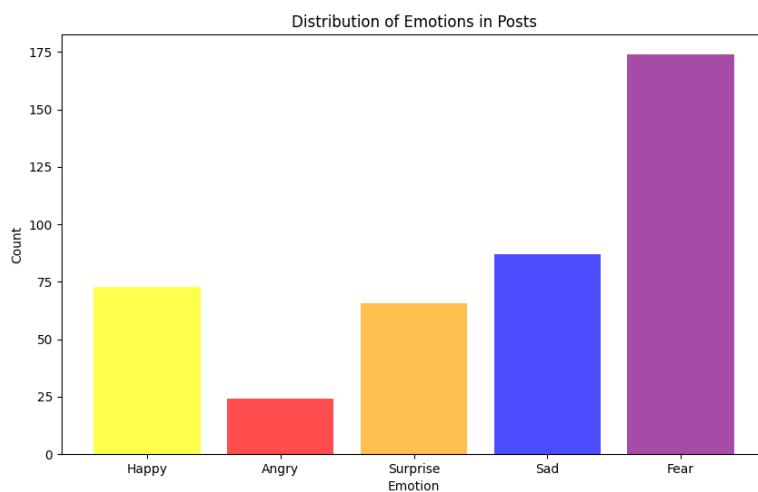
4.5.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie ponownie BERT ocenił więcej tekstów negatywnie (ponad 450). Postów ocenionych pozytywnie jest około 40.

4.5.5 Dystrybucja przypisanych emocji



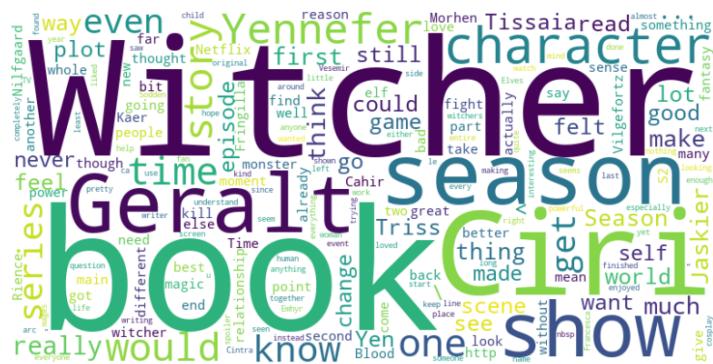
Dystrybucja emocji jest bardzo podobna do poprzedniej, jednak smutku ślupek znacznie zmalał. Więzkość postów ma dalej strachliwy wydźwięk, jednak niewiele jest złości.

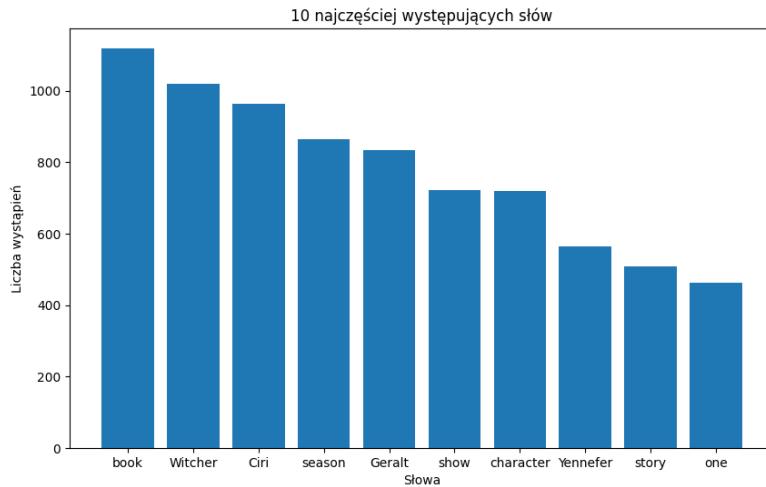
4.6 12.2021 - 06.2022

Na początku tego okresu ma premierę drugi sezon serialu i można zobaczyć, że liczba postów dostępnych w tym okresie gwałtownie wzrosła.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 200763, a po przeprowadzeniu lematyzacji - 84615.

4.6.1 Bag of words i Word cloud

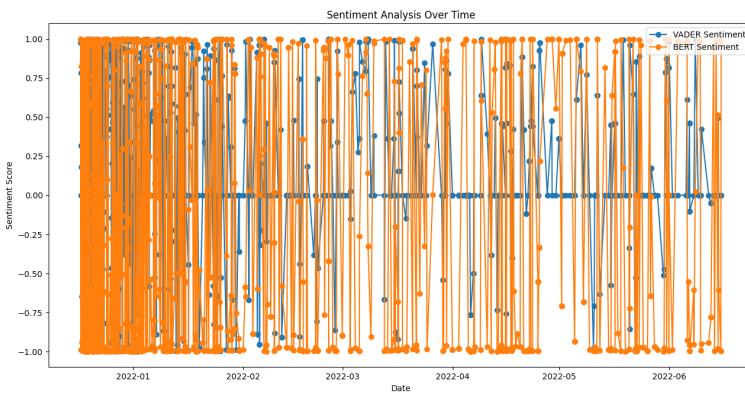




W tym półroczu tym razem na pierwsze miejsce wróciło słowo *book*, spychając wyraz *Witcher* na drugie miejsce. W tym zestawieniu można zobaczyć też słowo *season* i *show* dosyć wysoko, co może znaczyć, że ludzie porównują te dwie odsłony. Można też założyć, że o książkach wiedźmińskich mówi się dużo więcej niż o premierze.

4.6.2 Analiza odbioru w określonym czasie

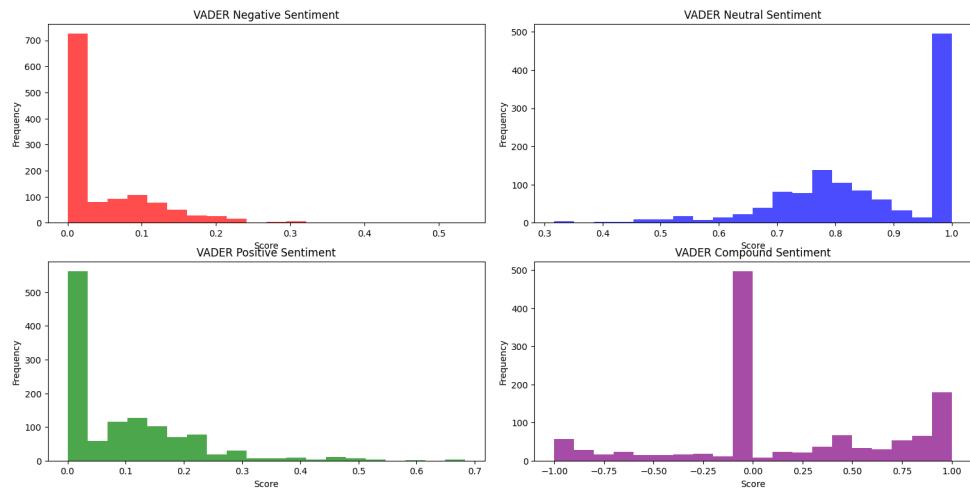
Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Jak widać jest bardzo duże zagęszczenie postów na koniec grudnia i przez cały styczeń, jednak później równomiernie spada. W porównaniu do okresu po wyjściu pierwszego sezonu, późniejsze zagęszczenie jest dużo mniejsze. Teksty dalej mają wydźwięk mieszany i, tak samo jak wcześniej, wyniki BERT

sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny. Choć są też przypadki, gdzie VADER jest zdecydowany o pozytywnym nastawieniu postu lub jego negatywnym nastawieniu.

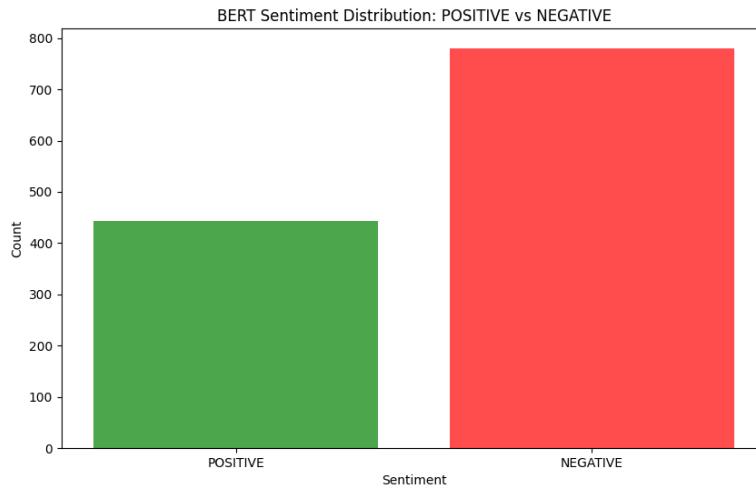
4.6.3 Analiza wyników przetworzonych przez VADER'a



Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach. Można tylko zwrócić uwagę na zmiany w detalach, gdzie pojawiło się całkiem sporo przypadków w pozytywnym histogramie większych niż 0.1. Tak samo są też negatywne komponenty większe od 0.1. Komponenty neutralne są dalej najbardziej zdecydowane, ale słupki mniej pewnych przypadków o argumentach 0.7 lub 0.8 znowu wzrosły.

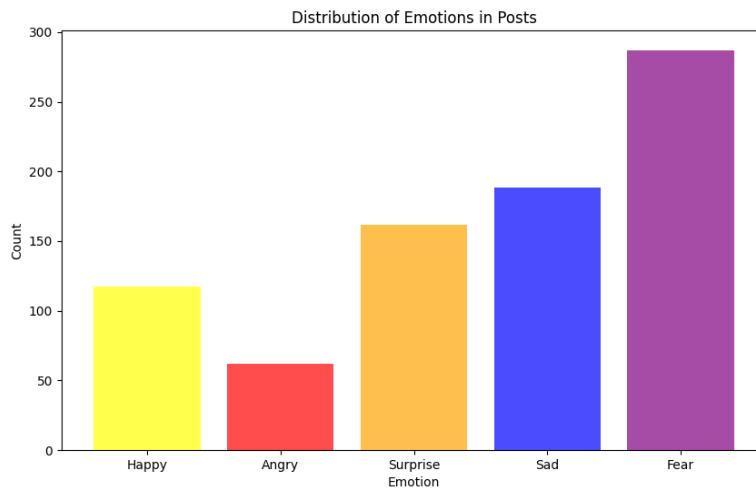
4.6.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie ponownie BERT ocenił znacznie więcej tekstów negatywnie (około 800). Postów ocenionych pozytywnie jest o połowę mniej (około 450).

4.6.5 Dystrybucja przypisanych emocji



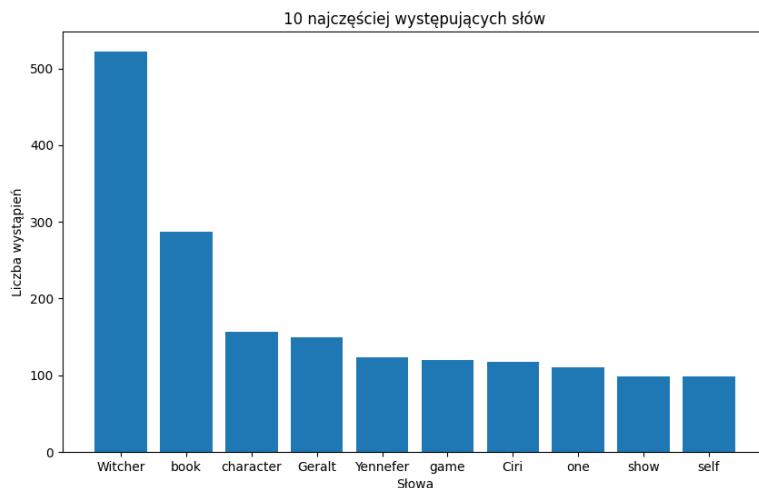
Dystrybucja emocji jest bardzo podobna do poprzedniej. Najwięcej niezeroowych wartości miał współczynnik strachu, później smutku, następnie emocji i szczęścia. Najmniej algorytm odczytał złości.

4.7 06.2022 - 12.2022

Teraz liczba postów znowu maleje, a przy serii zostają tylko najwierniejsi fani.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 41947, a po przeprowadzeniu lematyzacji - 18588.

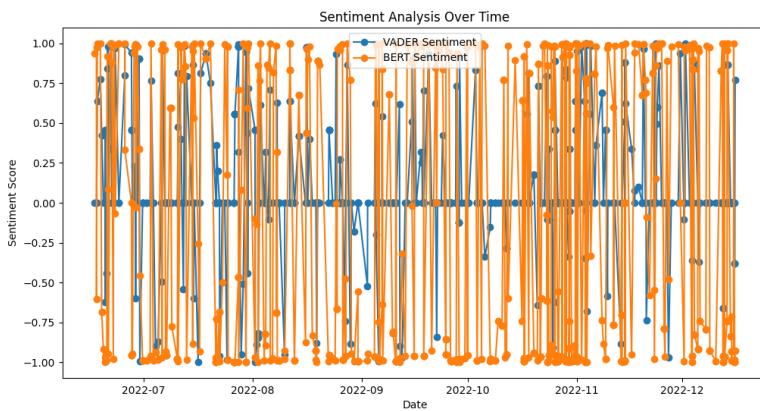
4.7.1 Bag of words i Word cloud



W tym półroczu na pierwsze miejsce wróciło słowo *Witcher*., spychając wyraz *book* na drugie miejsce. Słowo *Witcher* pojawiło się aż dwukrotnie więcej razy, niż wyraz *book*. W tym zestawieniu można zobaczyć też słowo *game* dość wysoko, co może znaczyć, że ludzie też odwołują się do gier z serii.

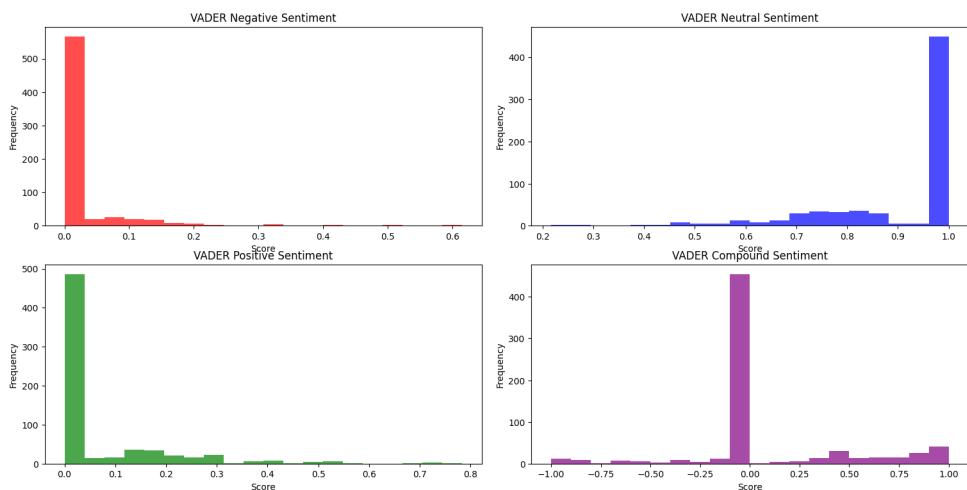
4.7.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Jak widać rozmieszczenie postów na wykresie jest w miarę równomierne. Jedyne zagęszczenie znajduje się pod koniec października i na początku listopada. Co ciekawe, jest to okres, w którym został zmieniony aktor grający postać Geralta w serialu Netflix'a i zostało to ogłoszone światu. Dalej nie dominuje tutaj ani wydźwięk pozytywny, ani negatywny. Posty dalej mają nastawienie mieszane i, tak samo jak wcześniej, wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny. Choć są też przypadki, gdzie VADER jest zdecydowany o pozytywnym nastawieniu postu lub jego negatywnym nastawieniu, jest to dosyć rzadkie.

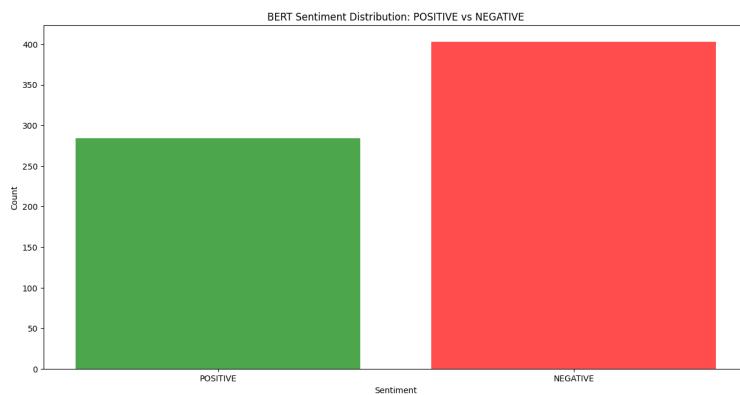
4.7.3 Analiza wyników przetworzonych przez VADER'a



Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach. Można tylko zwrócić uwagę na zmiany w detaluach, gdzie zniknęło większość bardziej zdecydowanych przypadków (większych niż 0.1) w pozytywnym histogramie, tak samo jak w negatywnym. Komponenty neutralne są dalej najbardziej zdecydowane, a słupki mniej pewnych przypadków o argumentach 0.7 lub 0.8 znów zmalały.

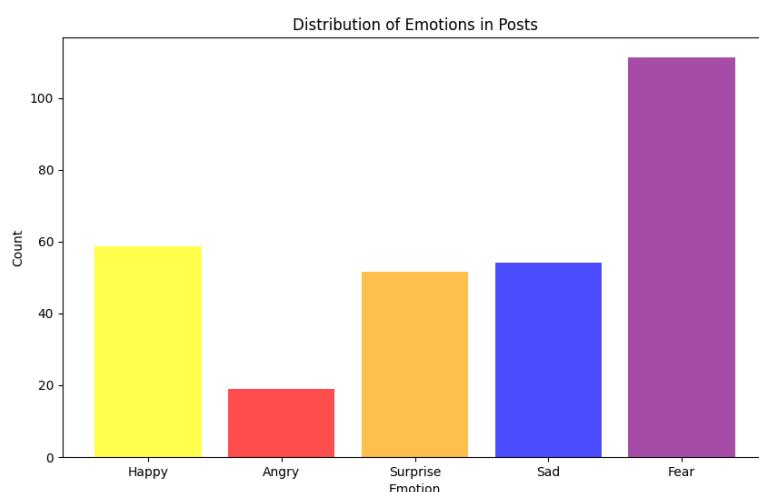
4.7.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie ponownie BERT ocenił więcej tekstów negatywnie (około 400). Postów ocenionych pozytywnie jest tylko około 275.

4.7.5 Dystrybucja przypisanych emocji



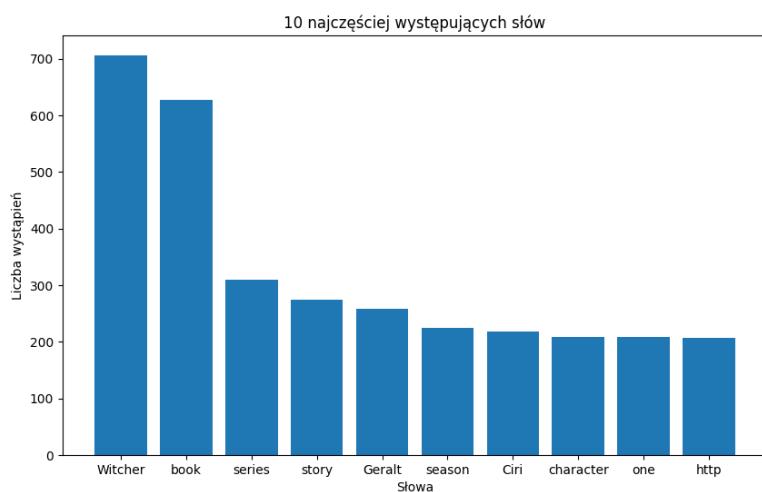
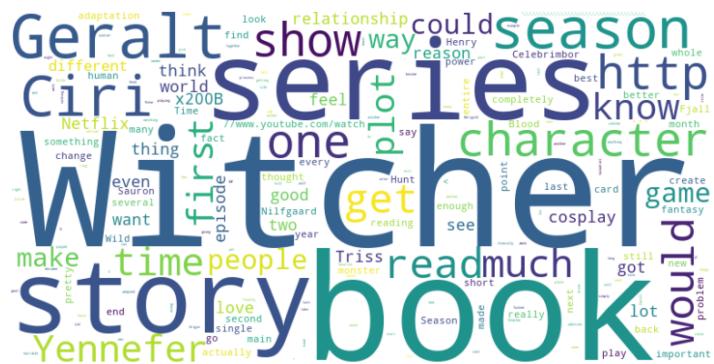
Co ciekawe, tym razem słupek z wykrytą emocją szczęścia zajmuje drugie miejsce pod względem wielkości, nie smutek.

4.8 12.2022 - 06.2023

Tym razem liczba postów jest troszkę większa zważywszy na fakt, że zbliża się premiera 3 sezonu serialu

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 83069, a po przeprowadzeniu lematyzacji - 34845.

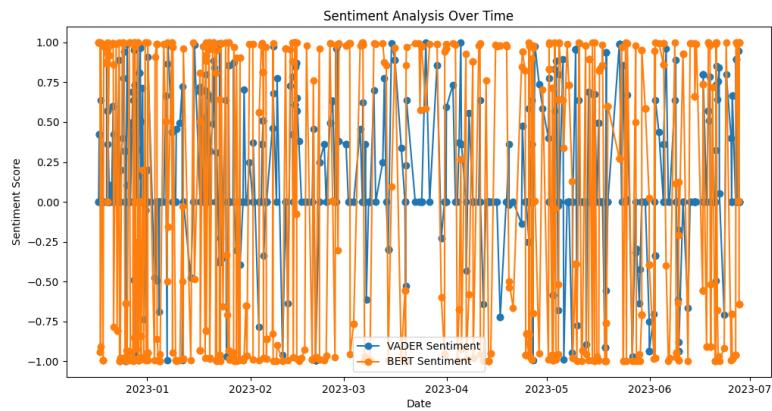
4.8.1 Bag of words i Word cloud



W tym półroczu na pierwszym miejscu jest słowo *Witcher*, jednak słowo *book* pojawił się niewiele mniej razy. W tym zestawieniu można zobaczyć też słowo *season* dość wysoko, co może znaczyć, że ludzi zaczyna też interesować kolejny sezon.

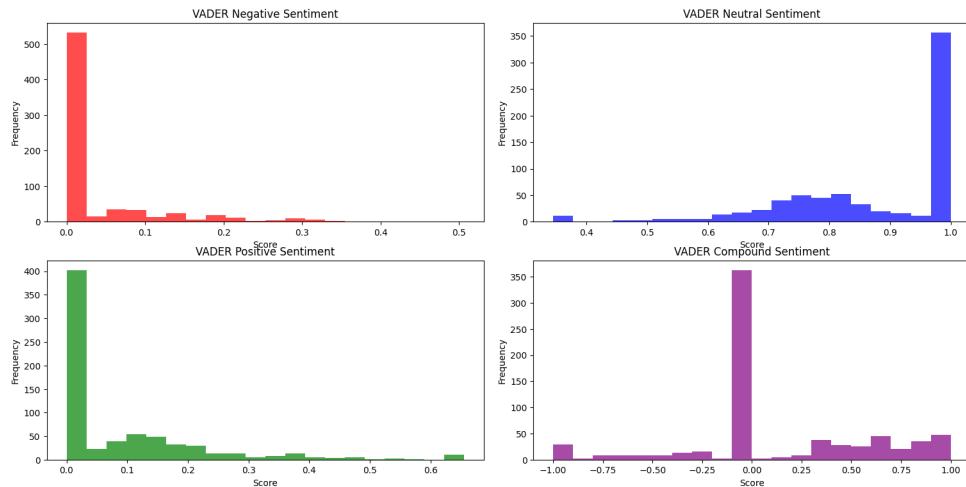
4.8.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Jak widać jest spore zagęszczenie postów w okolicach przełomu lat 2022 - 2023. Następnie trochę maleje, po czym znowu zaczyna rosnąć, jak zbliża się premiera 3 sezonu serialu w czerwcu. Jednak nie ma już tak wielu zainteresowanych, jak podczas premiery pierwszego sezonu. Dalej nie dominuje tutaj ani wydźwięk pozytywny, ani negatywny. Posty dalej mają nastawienie mieszane i, tak samo jak wcześniej, wyniki BERT sentiment są bardziej konkretne, gdzie wyniki VADER'a wskazują na wydźwięk neutralny.

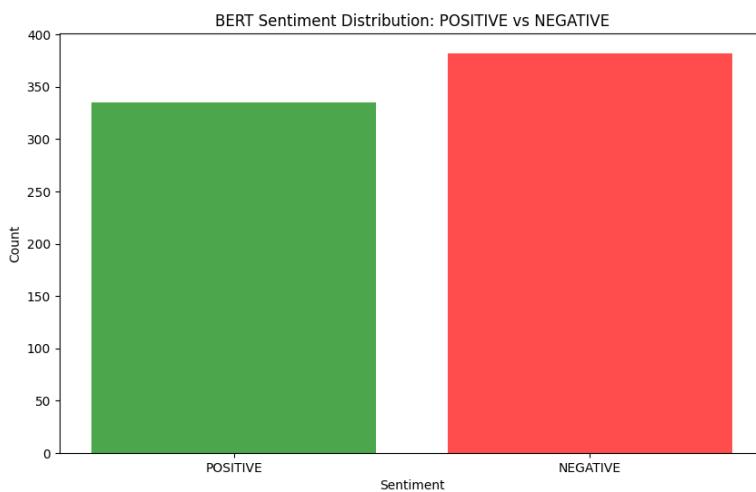
4.8.3 Analiza wyników przetworzonych przez VADER'a



Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach. Komponenty neutralne są dalej najbardziej zdecydowane, ale słupki mniej pewnych przypadków o argumentach 0.7 lub 0.8 tym razem znowu wzrosły.

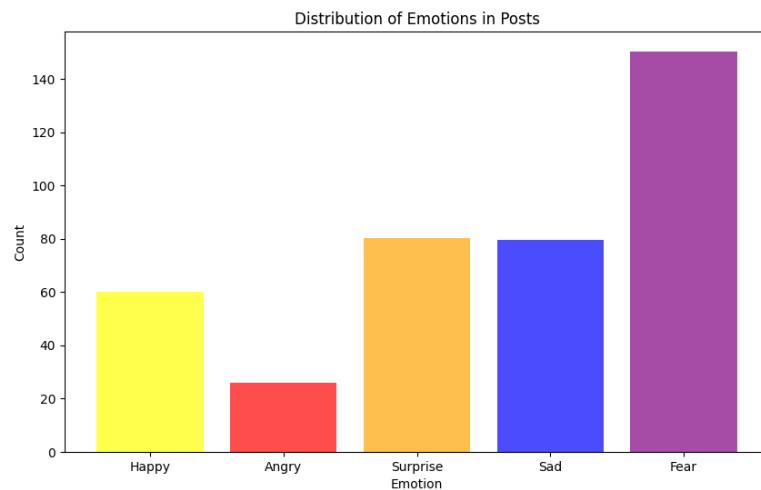
4.8.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie BERT ocenił nieco więcej postów negatywnie niż pozytywnie.

4.8.5 Dystrybucja przypisanych emocji



Teraz słupki z wykrytą emocją smutku i zaskoczenia się zrównały pod względem wielkości, wyprzedzając słupek emocji szczęścia.

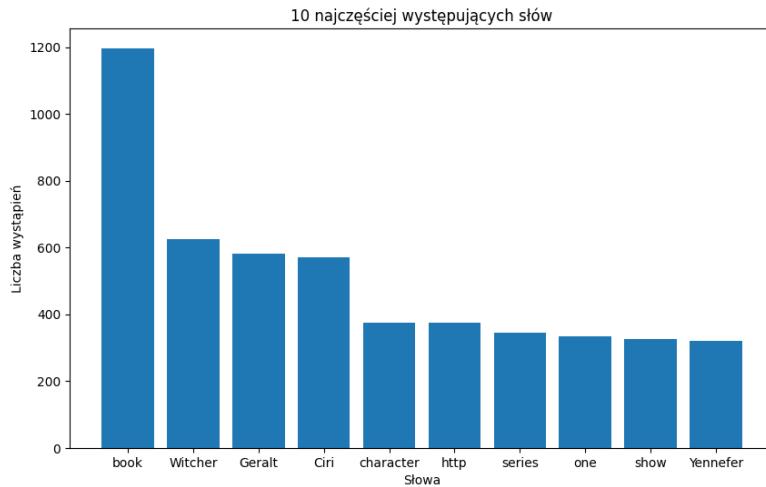
4.9 06.2023 - 12.2023

Jest to okres po wyjściu trzeciego sezonu.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 136258, a po przeprowadzeniu lematyzacji - 57781.

4.9.1 Bag of words i Word cloud

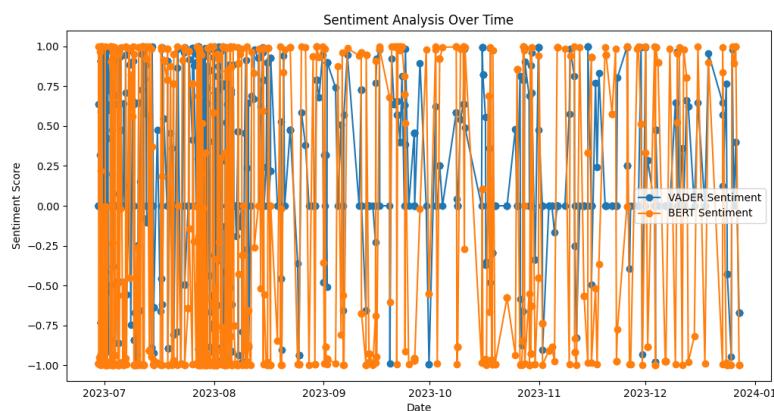




W tym półroczu dominuje słowo *book*, co znaczy, że wiele osób skupia swoje wypowiedzi wokół książek napisanych przez Andrzeja Sapkowskiego.

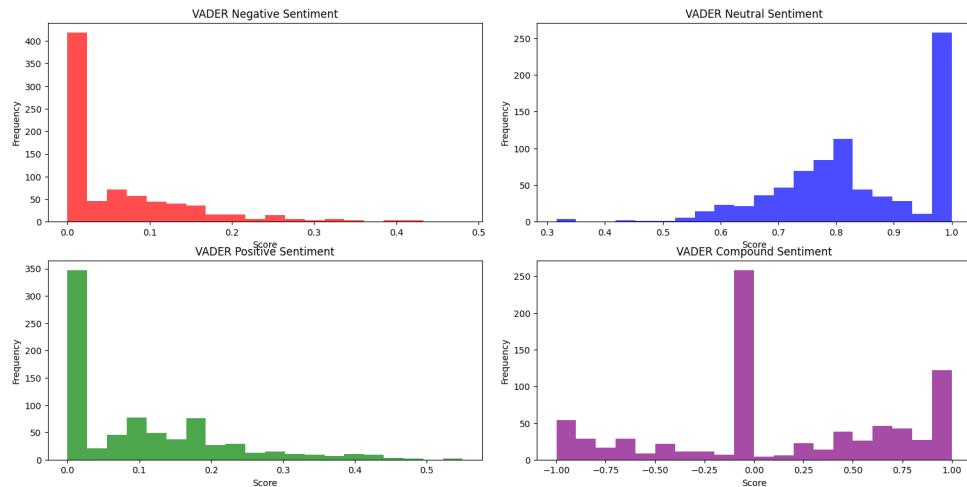
4.9.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Jak widać jest spore zagęszczenie postów w okolicach czerwca, lipca i sierpnia tego roku. To może być spowodowane premierą 3 sezonu, która została rozbita na dwa terminy. Pierwszy był 28 czerwca 2023 roku, a drugi w lipcu. Nie można też powiedzieć na podstawie wyników tego wykresu, czy większość wypowiedzi była pozytywna, czy negatywna. Ogólna opinia wydaje się być mieszana.

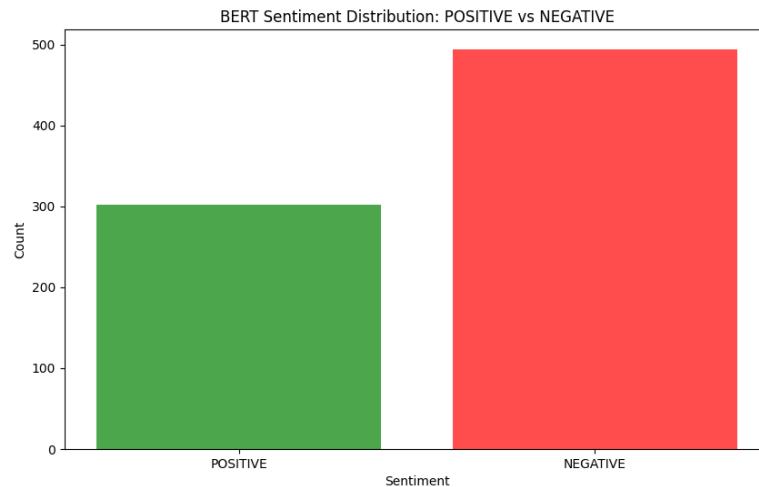
4.9.3 Analiza wyników przetworzonych przez VADER'a



Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach. Komponenty neutralne są dalej najbardziej zdecydowane. Do tego słupki mniej pewnych przypadków o argumentach 0.7 lub 0.8 tym razem wzrosły jeszcze bardziej.

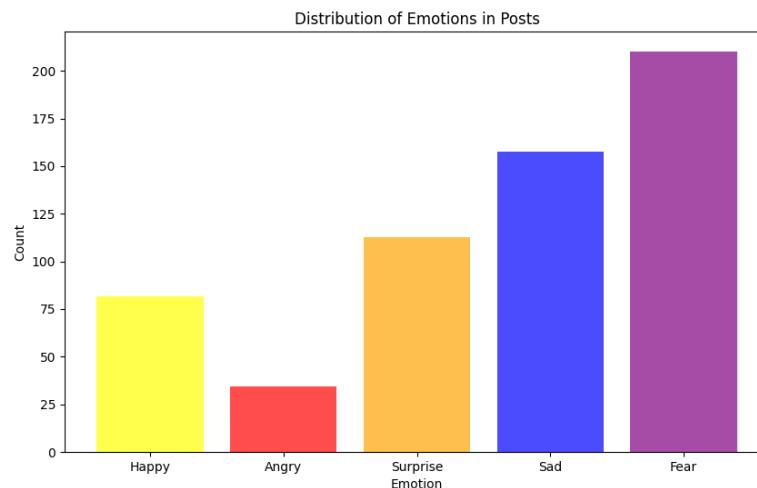
4.9.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie BERT ocenił prawie dwukrotnie więcej postów negatywnie niż pozytywnie.

4.9.5 Dystrybucja przypisanych emocji



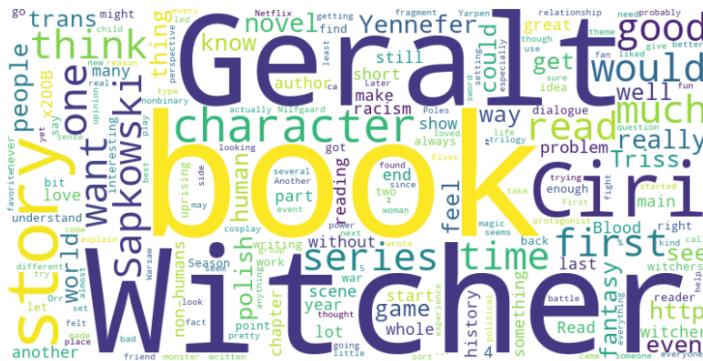
Dystrybucja emocji jest bardzo podobna do poprzedniej. Najwięcej niezeroowych wartości miał współczynnik strachu, później smutku, następnie emocji zaskoczenia i szczęścia. Najmniej algorytm odczytał złości.

4.10 12.2023 - 06.2024

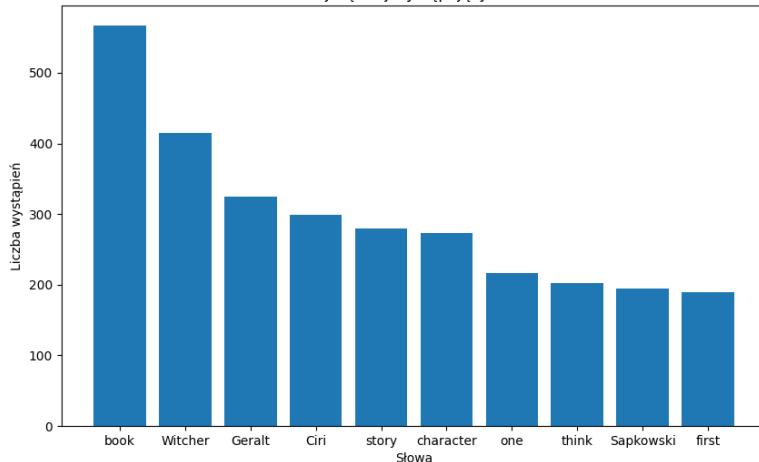
Jest to czas, w którym to opracowanie jest pisane. Nie ma żadnych innych danych z kolejnego okresu na dzień dzisiejszy.

Po tokenizacji łączna liczba słów we wszystkich postach z tego okresu wynosi 96933, a po przeprowadzeniu lematyzacji - 42152.

4.10.1 Bag of words i Word cloud



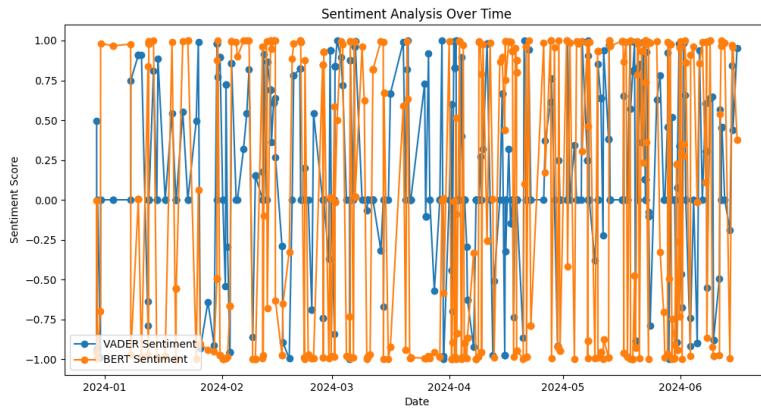
10 najczęściej występujących słów



W tym półroczu dominuje słowo *book*, co znaczy, że wiele osób skupia swoje wypowiedzi wokół książek napisanych przez Andrzeja Sapkowskiego.

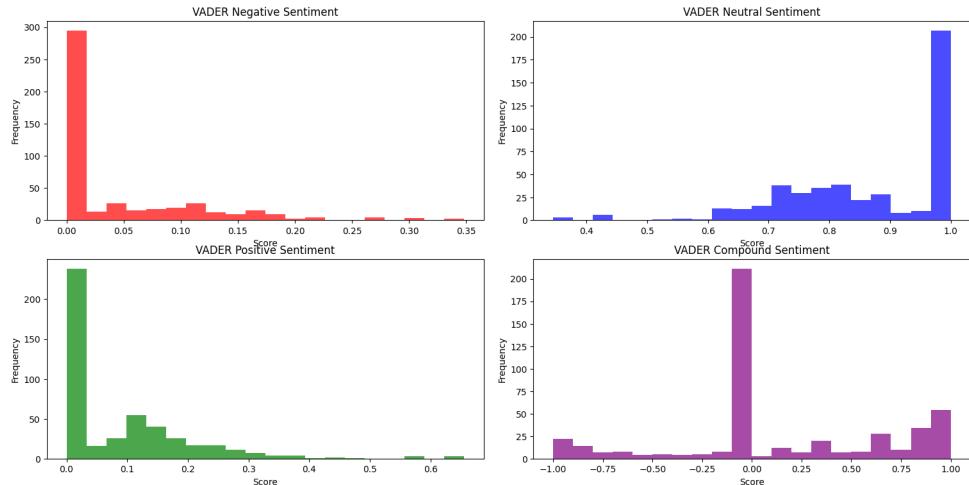
4.10.2 Analiza odbioru w określonym czasie

Poniżej pokazany jest wykres pokazujący zagęszczenie postów w danym czasie, a także ich nastawienie, pozytywne (wartość 1.00) lub negatywne (wartość -1) w danym okresie czasu.



Jak widać w tym okresie nie ma zbyt dużego zagęszczenia. Jest to czas dosyć spokojny. Opinie są dalej mieszane.

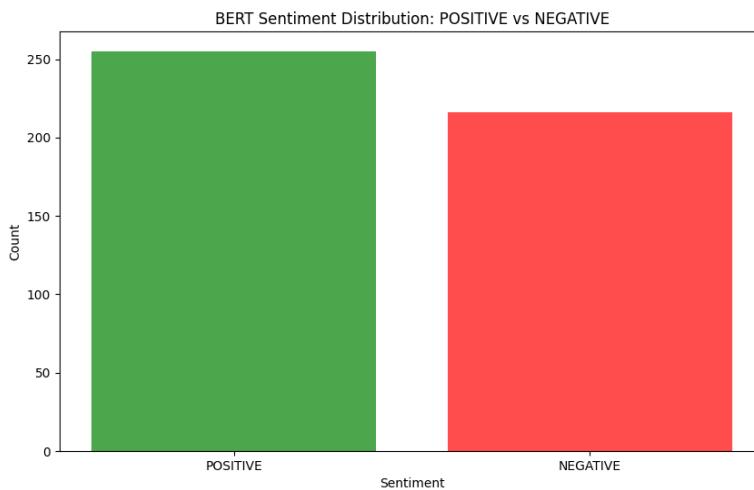
4.10.3 Analiza wyników przetworzonych przez VADER'a



Znowu, rozłożenie VADER'a jest podobne, co w poprzednich okresach. Komponenty neutralne są dalej najbardziej zdecydowane. Do tego słupki mniej pewnych przypadków o argumentach 0.7 lub 0.8 tym razem zmalały. Natomiast liczba większych współczynników w pozytywnym histogramie wzrosła.

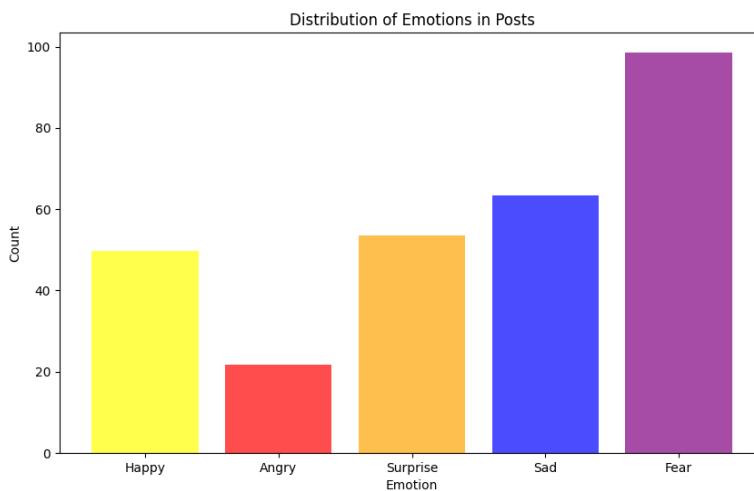
4.10.4 Analiza wyników przetworzonych przez BERT'a

Histogram poniżej zestawia ilość postów uznanych za negatywne i pozytywne.



W tym okresie BERT ocenił o wiele więcej postów pozytywnie.

4.10.5 Dystrybucja przypisanych emocji



Dystrybucja emocji jest bardzo podobna do poprzednich. Najwięcej niezerowych wartości miał współczynnik strachu, później smutku, następnie emocji i szczęścia. Najmniej algorytm odczytał złości.

5 Podsumowanie

Można zauważyc, 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 potórzyło, jednak zostali jej lojalni najwięksi fani. Dzięki temu seria w najbardziej spokojnych okresach była głównie chwalona, nie krytykowana.