

Sprawozdanie

Analiza Sentymentu Opinii o Filmie "Igrzyska Śmierci: Ballada ptaków i węży" za pomocą Biblioteki NLTK w pythonie.

Wprowadzenie

W ramach zadania przeprowadzono analizę sentymentu opinii na temat filmu "Igrzyska Śmierci: Ballada ptaków i węży". Wykorzystano do tego celu Pythona i bibliotekę Natural Language Toolkit (NLTK), która jest standardowym narzędziem do przetwarzania języka naturalnego.

Metodologia

Opinie zostały pobrane z serwisu Rotten Tomatoes:

https://www.rottentomatoes.com/m/the_hunger_games_the_ballad_of_songbirds_and_snakes/reviews?type=verified_audience

Następnie użyto następujących funkcji z NLTK:

- SentimentIntensityAnalyzer z modułu nltk.sentiment do oceny sentymentu każdej opinii.
- Tokenizacji tekstu i usuwania słów nieistotnych (stopwords) nie były wymagane, ponieważ SentimentIntensityAnalyzer automatycznie przetwarza tekst.

Proces Pobierania i Analizy Opinii

Użyto funkcji `get_reviews(url)` do pobrania opinii z podanej strony. Funkcja ta skrapowała tekst recenzji z wykorzystaniem bibliotek `requests` i `BeautifulSoup`.

Każda pobrana opinia była przekazywana do SentimentIntensityAnalyzer z NLTK, który analizował sentyment tekstu, zwracając wyniki w postaci słownika z wartościami dla 'neg', 'neu', 'pos' oraz 'compound'.

Wyniki i ich interpretacja

Wyniki 'neg', 'neu', 'pos' reprezentowały procentowy udział odpowiednio negatywnych, neutralnych i pozytywnych emocji w tekście. Wartość 'compound' stanowiła znormalizowaną ogólną ocenę sentymentu tekstu, waha się od -1 (bardzo negatywny) do +1 (bardzo pozytywny). Oto przykłady interpretacji kilku wyników:

Recenzja 1: *"It's nothing I didn't dislike I loved it"*

- **Sentiment:** {'neg': 0.337, 'neu': 0.429, 'pos': 0.234, 'compound': -0.2411}
- **Interpretacja sentymentu:** Mieszany sentyment z przewagą negatywnych emocji, co może wynikać z podwójnego zaprzeczenia użytego w recenzji.
- **Rzeczywiste Odczucie:** Recenzja wyraża pozytywną opinię ("I loved it"), jednak skomplikowana struktura zdania wprowadza zamieszanie w analizie sentymentu.

Recenzja 2: *"Good adaptation of the book. Missing some detail but you can't make a 14hr movie."*

- **Sentiment:** {'neg': 0.103, 'neu': 0.772, 'pos': 0.125, 'compound': 0.09}
- **Interpretacja sentymentu:** Lekko pozytywny, ale w większości neutralny sentyment, odzwierciedlający mieszane uczucia recenzenta.
- **Rzeczywiste Odczucie:** Recenzja wyraża ogólnie pozytywną opinię o adaptacji, ale zauważa braki, co jest dobrze odzwierciedlone w wyważonym wyniku sentymentu.

Recenzja 3: *"How many times can they say her name"*

- **Sentiment:** {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
- **Interpretacja sentymentu:** Neutralny sentyment, sugerujący brak wyraźnych emocji.

- **Rzeczywiste Odczucie:** Recenzja może mieć lekko negatywny wydźwięk, sugerujący irytację, ale jest zbyt niejasna, aby ją jednoznacznie sklasyfikować, co może tłumaczyć neutralny wynik.

Recenzja 4: *"Should've been split into two movies"*

- **Sentyment:** {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
- **Interpretacja sentymentu:** Neutralny sentyment, brak wyraźnych emocji.
- **Rzeczywiste Odczucie:** Recenzja sugeruje pewien negatywny aspekt (przeciążenie treści), ale ogólnie jest zbyt ogólna, aby wyraźnie określić sentyment.

Recenzja 5: *"Really good acting makes this movie a good one! Viola Davis was amazing as was the actor who played Snow. The snakes were a little fake but it worked."*

- **Sentyment:** {'neg': 0.059, 'neu': 0.677, 'pos': 0.264, 'compound': 0.689}
- **Interpretacja sentymentu:** Głównie pozytywny sentyment, z niewielką negatywnością odnoszącą się do "fake snakes".
- **Rzeczywiste Odczucie:** Recenzja jest wyraźnie pozytywna z drobną krytyką, co jest adekwatnie odzwierciedlone w wynikach sentymentu.

Recenzja 6: *"Absolutely Mind Blown"*

- **Sentyment:** {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
- **Interpretacja sentymentu:** Neutralny sentyment, co może być związane z brakiem wystarczających informacji w tekście do analizy.
- **Rzeczywiste Odczucie:** Ta recenzja jest wyraźnie pozytywna, co nie zostało wykryte przez analizę sentymentu. Pokazuje to ograniczenie w rozpoznawaniu krótkich, ale intensywnie pozytywnych opinii.

Wnioski

Analiza sentymentu wykonana za pomocą biblioteki NLTK w Pythonie na recenzjach filmu "Igrzyska Śmierci: Ballada ptaków i węży" z serwisu Rotten Tomatoes dostarcza cennych wglądów w ogólne nastroje i emocje wyrażane przez recenzentów. Jednakże, wyniki tej analizy pokazują, że automatyczne narzędzia do analizy sentymentu mogą mieć ograniczenia, szczególnie w kontekście złożonych struktur językowych, sarkazmu, ironii czy subtelnych niuansów emocjonalnych.

W przypadkach, gdy recenzje zawierają proste i bezpośrednie wyrażenia emocji, analiza sentymentu NLTK wykazuje dobre dopasowanie do rzeczywistych odczuć. Jednak w sytuacjach, gdy tekst jest bardziej złożony lub zawiera elementy mogące wprowadzać w błąd (jak podwójne zaprzeczenia), wyniki mogą być niejasne lub nieprecyzyjne. To wskazuje na to, że choć narzędzia takie jak NLTK są pomocne w analizie dużych zbiorów danych tekstowych, powinny być używane z ostrożnością i najlepiej jako uzupełnienie ludzkiej interpretacji.

Podsumowując, analiza sentymentu za pomocą NLTK oferuje przydatny wgląd w ogólne nastroje i tendencje w tekstach, ale wymaga dokładnego zrozumienia jej ograniczeń oraz potencjalnego uzupełnienia przez interpretację ludzką, zwłaszcza w przypadku subtelnych lub skomplikowanych wyrażeń emocjonalnych.