

Analiza opinii publicznej na podstawie wpisów z Twittera(Python)

Marek Wiecaszek

June 15, 2022

1 Wstęp

Kozystając z biblioteki nltk dokonamy predykcji opinii publicznej Johnego Deppa oraz Amber Heard przed 2016 w okresie przejściowym oraz stanem obecnym. Zakres czasowy został wybrany na podstawie kontrowersyjnych momentów z życia aktorów, które u wielu wywołują zmienne uczucia. Do analizy zostało użyte 30 tys. tweetów z przed 2016, po 15 tys. dla każdej z osób, 500 z każdego miesiąca z okresu przejściowego, oraz kolejne 30 tys. jak najnowszych wpisów. Wydaje mi się, że jest to wystarczająca ilość danych do przeprowadzenia analizy.

2 Predykcja przed 2016

2.1 Uwagi

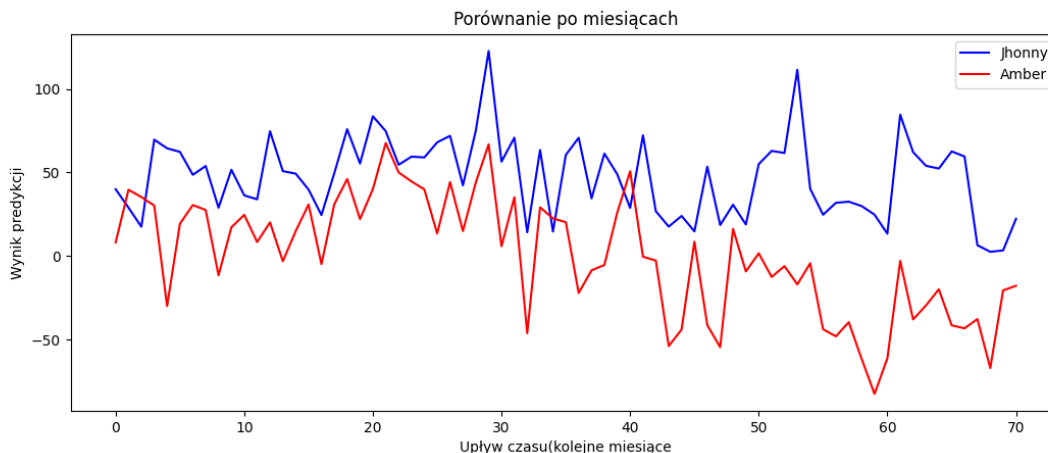
Oprócz wyników funkcji `SentimentIntensityAnalyzer` wbudowanej w moduł nltk wykonałem tzw. `wordcloud`'y, aby lepiej zobrazować jakie słowa najczęściej były używane przez użytkowników w połączeniu z hashtagami `johnnydepp` i `amberheard`. Aby miały one jakikolwiek sens usunięte zostały z nich tzw. `stopwordy` czyli zaimki, linki to stron internetowych i inne mało znaczące słowa mogące zniekształcać wynik. Słowa zostały również poddane leminizacji czyli ujednoliceniu odmiany. W ten sposób słowa takie jak `loved` czy `loving` zostały zamienione na podstawową wersję `love`. Oczywiście do samej analizy uczuć tweety nie były poddawane tym operacjom.

2.2 Johnny Depp

Funkcja odpowiedzialna za predykcje zwraca 4 podstawowe wyniki. Jest to ocena pozytywna, neutralna, negatywna oraz ocena zbiorcza (ang. `compound`), która jest najważniejsza. W przypadku Johnego Deppa Uzyskał on wynik `compound` na poziomie 2308, który powstał po zsumowaniu ze sobą wyników poszczególnych tweetów. Biorąc również pod uwagę różnice w wynikach pozytywnych i negatywnych 2550 i 850 kolejno oznacza to, że wypowiedzi ludzi na jego temat było dość pozytywne. Można to również wywnioskować z poniższego obrazka, który przedstawia najczęściej używane przez nich słowa z "love" na czele.

3 Okres przejściowy

Aby zobrazować jak opinia publiczna na temat dwójki aktorów zmieniała się pomiędzy 2016, a 2022 został przygotowany wykres, który przedstawia wyniki funkcji predykcji dla każdego miesiąca.



Jak widać pomimo dość dużej dyzaprobaty Johnnego przez big media po ogłoszeniu przez Amber, że była ofiarą przemocy oraz wzięciu rozwodu przez te dwójke opinia publiczna przez cały okres była bardziej przychylna dla pana Deppa. Można również zaobserwować dość silne powiązanie ze sobą wyników co jest spowodowane tym, że spora część wpisów pokrywała się ze sobą (zawierała oba wcześniej wspomniane hashtagi). Dodatkowo wraz z upływem czasu i ujawnianiem dalszych informacji na temat tego co wydarzyło się naprawdę różnica pomiędzy wynikami stawała się coraz większa.

4 Stan obecny

Aktualne wyniki przedstawiają stan po ogłoszeniu wyników rozprawy sądowej i zwycięstwie Johnnego Deppa. Biorąc pod uwagę co działo się przez ostatnie kilka tygodni nie są one zaskakujące.

4.1 Johnny Depp

Wyniki aktora znanego z roli kapitana Jacka Sparrowa są jeszcze lepsze niż były wcześniej. Uzyskał on bowiem wynik compound 3616. Znacząco też wzrósł wynik pozytywny i negatywny (4703 i 2108). Tak wysoki wynik negatywny jest prawdopodobnie spowodowany tym, że johnnydepp pojawiał się często we wpisach o Amber. Na poniższym obrazku widać, że hasła najczęściej występujące w tych wpisach to "love", "happy birthday", czy dość obraźliwe "ambert**d".



4.2 Amber Heard

Aktorka dużo straciła pozywając swojego byłego męża. Nie tylko musi zapłacić mu odszkodowanie, ale również jej opinia publiczna legła w gruzach. Wynik compound dla wpisów na jej temat to - 1831 (pozytywne 2834, negatywne 3673). Dodatkowo najczęściej występujące słowo to "lie". Można również zobaczyć słowa "abuser" czy "abuse". Pojawiają się również zwroty pozytywne jak "believe", jednak znając kontekst możemy je łatwo przypisać do słów aktorki, które parafrazując brzmiały mniej więcej tak "Go tell the World Johnny and see if they believe you".



5 Podsumowanie

Jak widać w dzisiejszych czasach można wykorzystywać komputery nawet do analizy ludzkich uczuć. Oczywiście jest to dość powierzchowne i aby poznać opinie publiczną na jakiś konkretny temat potrzeba tysięcy indywidualnych wypowiedzi. Myślę jednak, że wyniki obliczone przez program pokrywają się w dużym stopniu ze stanem rzeczywistym co jest niesamowite.

References

<https://www.nltk.org/howto/sentiment.html>

<https://www.datacamp.com/tutorial/wordcloud-python>

<https://twitter.com>