Working with Tweets

Nanda Rajiv
2021115002

Twitter Hashtags Scraped – ["#djocovid", "#djokovic", "#novaxdjocovid","#novaxdjokovic","#nolefam", "#westandwithnovak", "#istandwithnovak", "#djokovicout", "#boycottaustralianopen"]

Using the above hashtags to scrape twitter and then using the Bag of Words program gives us a bird's eye view of what the narrative of the controversy around Novak Djokovic and the 2022 Australian Open was. Separately using Bag of Words on each hashtag can also give us an idea of the language and buzz words that are most often used by each side of this controversy.

Before looking into the results of the programs, a general of timeline of the events surrounding the issue is established. In April 2020, Novak Djokovic publicly opposes vaccine mandates, and spurs widespread criticism. On 4 Jan 2022, Novak Djokovic posts an image on social media wishing his fans a happy new year, and informing that he is heading down to Australia for the Australian Open, with a medical exemption granting him entry to the country without having taken the vaccine. At the time, entry with valid medical exemption was permitted in special cases. He lands the next day in Melbourne, his visa is cancelled, and he is detained at a state-run immigration facility. His parents criticize the move, and the Serbian president issues a statement criticizing Australian PM Morrison. After a few days, a court overturns his visa cancellation, and Djokovic starts training. He is included in the Australian Open draw as the first seeded player. On Jan 14, his visa is cancelled for a second time. Djokovic's second legal appeal is rejected, and he is deported on Jan 16.

During the entire course of events, several people spoke out publicly both in support of and criticizing Novak Djokovic, for holding his stance against vaccine mandates.
The Australian government is also praised and alternately criticized for their changing positions on whether Djokovic's entry was permitted or not.

Hashtags in support of Djokovic – ["#nolefam", "#westandwithnovak", "#istandwithnovak", "#boycottaustralianopen"]

The Bag of Words run on these individual hashtags yielded the most common words as 'Djokovic' and 'Novak', as well as their respective hashtags, which is expected, since all tweets would contain these. Apart from these, words like 'freedom', 'choice' and 'tyrannical' are common, which aligns with the general narrative of people not aligned with vaccine mandates. The #NoleFam (Nole being Djokovic's nickname) Bag of Words did not engage with the vaccine controversy as much as the others, with words like 'fans' or 'goat' (common tennis abbreviation for 'greatest of all time') being the most used.

Hashtags in criticism of Djokovic – ["#djocovid", "#djokovicgohome", "#novaxdjokovic", "#novaxdjocovid"]

Here too, the top most common words are 'Novak' and 'Djokovic', as expected. Versions of 'medicine' or 'medical' are common, as well as the word 'infection' and 'protocol', which were not there in the previous section. There is also some conjecture and debate on whether or not Djokovic will be allowed to participate in future tournaments – 'Roland-Garros' and 'Indian Wells' also making it on the Bag of Words results.

Neutral Hashtags – ["#djokovicout", "#djokovic"]

We see the least engagement with the vaccine controversy on these hashtags. Among the most common words are 'Nadal' and 'Medvedev', the players that were in the Australian Open Men's Final in the end. There is more speculation about future tournaments, including 'Wimbledon', 'Roland-Garros' and 'Indian Wells'. There have been recent reports that Djokovic is on the entry list for vaccinated-only Indian Wells tournament set to be held in March.

Overall, there is a difference in the lingo used across different hashtags, as well as the level of engagement with the issue. Though analyzing through this method of scraping tweets of a certain hashtag and running it through the Bag of Words program does give us a useful insight, it is not a foolproof way of identifying public perception on a certain issue. In addition to identifying the words used in tweets, it would also be useful to look into *who* is tweeting what – their nationality, political leaning, whether or not they actively follow the sport, etc.