

Ex 4: Newspaper Analysis

Arghya Roy
2021115008

Topic: Winter Olympics 2022

Newspaper: Times of India (<https://timesofindia.indiatimes.com/>)

After scraping the recent Times of India articles and exporting them to a file, I performed the bag-of-words exercise on it and analysed the results.

Analysis:

The most frequent notable word turns out to be “Beijing”. This is quite expected as the Winter Olympics are being held at Beijing this time. The word “Shailaja” is also frequent. This is because we know that India’s representation in the Winter Olympics has always been very sparse- in fact, Arif Khan, the sole Indian in this edition of the Winter Olympics, happens to be just the 16th Indian participant across all 11 editions of the Winter Olympics. So, in that context, the story of Shailaja Kumar is coming to the attention of many as she was the first female Indian ever at the Winter Olympics(1988).

The next word that popped up is one that has been causing quite the buzz all across the globe- “Valieva”. This is because Kamila Valieva is a 15-year-old Russian figure skater, whose doping test returned a positive result after the conclusion of a national event. She was initially served a ban which was soon lifted owing to the fact that she is a minor. She even went on to win a gold medal at the Winter Olympics. This controversy caught the attention of many. This word “Russian” is also very frequent, mainly owing to this scandal and also because Russia has the second-best total medal tally in the Winter Olympics. “Norway” is also very frequent as Norway actually have the best medal tally of all participating nations.

Words like “gold”, “silver”, “bronze”, “medal”, “competition”, “sport”, etc. are very frequent and expectedly so since the Winter Olympics are one of the biggest and most significant sporting competitions in the world. Also, words like “slalom” are “skiing” are frequent as they are some of the many sports that are a part of the event.

In conclusion, I feel that the bag-of-words exercise may not have been the most relevant computational tool for this instance. Most of the analysis provided needed extensive factual connecting-of-dots and just the results of the bag-of-words exercise itself didn’t provide much except for stray hints. The articles that were scraped were all common to the Winter Olympics. They are otherwise about mostly disjoint events. Instead, if we would have chosen particular sub-events of the Winter Olympics (say, the doping scandal mentioned before) and scraped the corresponding articles from various newspapers from different regions across the world, then it would have been much more relevant and riveting.

The relevant files(codes, outputs, URLs of the articles scraped etc.) are all attached. They can also be found [here](#).