Uncovering insights on kidnaps in Uganda using text mining

Abstract

Kidnaps are deeply traumatizing incidents to society and mostly to the families that experience them. While some victims may lose their lives[1], others are trafficked or go missing without a trace. In low-income countries like Uganda, where digitized data on incidents such as kidnaps is incoherent, it becomes hard to employ the use of data science techniques directly to the available data. Nonetheless, the currently available data is mainly through government bodies like the Uganda National Bureau of Standards(UBoS) and Police reports, which are also not easily accessible or even reliable. In this digital era, case investigations can be overwhelming when employing traditional data collection methods like physical interviews and face a challenge of under-utilization of information sources like social media sites. While text mining is used to uncover critical insights into political biases, voter turnout, election prediction, and economic indicators, it also has many times been used to understand criminal social behavior, trends, and patterns. This research work sought to gather a dataset from alternative sources that are representative of the situation on the ground and present an implementation of data gathered through text mining from both local news media(New Vision) website and Twitter, to try to uncover insights and trends of the kidnaps in the country on these platforms. In the work, we used Social media to collect data and analyze it to understand the trends in kidnaps and draw insights and patterns on what drives the incidents in the country.

Methodology and Results

Through web scraping the new vision website[1] and twitter for articles and tweets respectively, on incidents of kidnaps in Uganda from Ist January 2018 to 30th July 2019, we generated a dataset. On the new vision website, the main search word was 'kidnap' from whose search results, we narrowed down our time frames. On Twitter, we used the advanced Twitter search option which was less costly than the API and searched kidnaps in Uganda with the same time frame. We came up with web scraping scripts to collect 164 reports from NewVision and another to get 102 from Twitter after which the data was cleaned using Pandas. The data from the different sources were then visualized using Matplotlib and Word Cloud to get a graphical representation of the word frequency within the datasets. Words with high prominence provide a starting point for further investigation and a better understanding of the situation. With the datasets, the word clouds generated showed an emphasis on Kidnap, women, security, money, ransom, Kampala (the country's capital), United States, Police, Court, Law, Children as most frequent. This raised a question of how the above may be affecting/ contributing to the kidnap cases that arise in the country. Future work will involve getting a larger dataset by increasing the timelines for the searches, using other media sites and also digging deeper into the frequent words to determine how they affect kidnap rates in the county. This will zero down the kidnapping hotspots and identify the possible culprits, thus reducing the crime rate in-terms of kidnaps in Uganda.

References

[1] Child Sacrifice in Uganda: adequacy of existing legal provision By Walakira Eddy and Phillipa Bogere