**Identification of Medical Misinformation on Twitter – Executive Summary**

**Background:** Laws permitting the use of Cannabis for medical and recreational use in many states across the US has resulted in substantial tweets on Twitter propagating unverified information about health benefits of Cannabis. Not many studies have been done in recent times that shed light on how much information on Twitter regarding Cannabis health benefits is verified. Individuals can get influenced by these tweets which would cause repeated use of cannabis resulting in health implications.

**Objective:** The primary objective was to identify medical misinformation regarding the use of Cannabis for medicinal purposes on Twitter. We also aimed to classify all tweets making health claims into whether they were made by social bots or individuals. As we delved deep into our analysis of Bot behaviour, we realized a substantial chunk of unverified cannabis health claims were tweeted by Bots which had a promotional theme to the tweet which prompted us to explore the behaviour of these bots as well.

**Methods:** The package GetOldTweets3 was used in Python to collect 378,348 tweets having cannabis related terms between the months of November 2019 and March 2020. This data was pre processed using the NLTK package in Python to aid in the analysis. Further, a semantics-based approach was used to identify tweets making health claims from the larger set of tweets collected earlier. The semantics approach had 2 Bag of words – Medical Conditions and Treatment. The Medical Conditions bag of words had 963 words listing all possible medical conditions and this was sourced from the CDC website and Stanford Health website. The Treatment Effects bag of words had 300 words which had words which described effects like ‘reduces’ , ’decreases’ , ’helps’ etc.

Using the above semantics-based approach, there were 3490 tweets filtered from the larger set of 378,348 tweets which were making health claims regarding the use of cannabis. Using the package BotorNot, the number of social bots and non-bots was analysed on both – tweet level and user level. After identifying bots, various behaviour of bots were analysed. Finally, the 3490 tweets making health claims were classified into verified and unverified tweets using a classifier which used verified facts sourced from the CDC website.

**Results:** Out of the 378,348 cannabis related tweets initially collected, only 3490 tweets contained medical health claims regarding the use of Cannabis. Out of the 3490 medical claims regarding Cannabis use tweets 822 tweets were made by Social Bots. On analyzing the behavior of Bots, the Bot activity was highest on Mondays out of all weekdays and highest in the month of December and it was highest at late night. On average Bots tweeting 1 or 2 tweets per day contributed to 70% of tweets. 80% of the User Mentions (@Username) in the tweets made by Bots are Verified twitter accounts with more than 100k followers. Out of the 822 tweets, 590 tweets contained URLs. Of the 590 tweets that contained URLs, 344 unique URLs were observed. Grouping Bots by the unique URLs present in the tweets, we found four possible Network of Bots in our data. Out of the 2551 final unique tweets there were 697 Promotional Tweets and 1890 Non- Promotional Tweets. Out of the 2551 final unique tweets there were 30 tweets that contained Verified Facts regarding the use of Cannabis and 2557 tweets that contained Unverified Facts. Out of the 215 Promotional Tweets made by Bots there were 0 tweets that contained Verified Facts regarding the use of Cannabis

**Conclusion:** 1 % of tweets (3490 tweets) were mentioning Medical claims about using Cannabis out of 378,348 tweets mentioning Cannabis keywords. Out of 2588 unique tweets mentioning medical claims of using Cannabis, only 1.12 % are Verified Facts and all other are Misinformation (Unverified Facts) regarding Medical usage of Cannabis. Tweets made by Bots contributed to 23.5% of total 3490 tweets and remaining were made by non bot accounts. Also there are 2128 unique user accounts which made Health claims regarding the use of Cannabis out of which 12% of user accounts (255 user handles) were bots and remaining 1839 were non-bots accounts. Bot seems to be most active on Monday which contributed to 28 % of total 822 tweets made by bots. There is no significant difference in tweets made by Bots on other days of week. Also most of tweets made by bots are in the month of December contributing to 32% of total 822 Bot tweets. For other months frequency of tweets made by Bot accounts is similar. Bots activity is highest at night and 70 % of Bots make 1 or 2 tweets per day. 72% of Bots tweets have URL in them and there were 10 unique user mentions (@person tagging) and which have having more than 100k followers. It seems that bots which makes user mentions seems to tag celebrities. Also 27 % of total 3490 tweets are promotional, and all promotional tweets made by Bots are Misinformation regarding Health claims about use of Cannabis