Project Robinhood- Analyzing political hate speech on Twitter during the Indian Assembly Elections 2022.

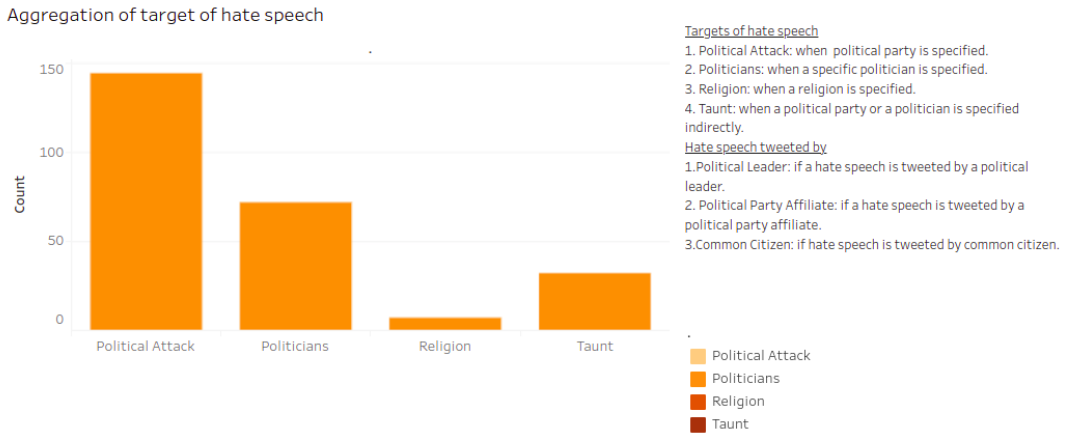
Come election season, and social media platforms become a hotbed of misinformation, hate speech, and other harmful content available in the form of textual or multimedia content. On the one hand, these platforms aid the political parties and citizenries to engage in democratic discourse regarding our electoral rights. On the other hand, these platforms are misused to spread harmful content that can be hateful, polarizing, and even instigate violence in the offline world. Around the time of the election, politicians leaders and their affiliates use social media platforms to share updates of their political campaigns, make announcements about contesting candidates and even criticize members of other competing parties. One can only imagine the quality and quantity of content posted by political parties and other politically active citizens to show varying patterns before, during, and immediately after the elections. To analyze such linguistic patterns from the point of hateful content, a team of researchers at the Laboratory for Computational, Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi) have recently launched Project Robinhood. The project will track content on the Twitter platform regarding Indian Assembly Elections 2022, curating information from January to March. As the analysis revolves around political hatred and hostility, to being with, the team is focused on collecting data of political leaders. Starting with 50 political leaders from 5 major parties (BJP, INC, BSP, SP, and AAP). To share the insights with a broader audience, the team has launched a portal that provides aggregated and granular statistics of the various groups of users generating political hate and the groups it is targeted against. This information is further analyzed at the State level.

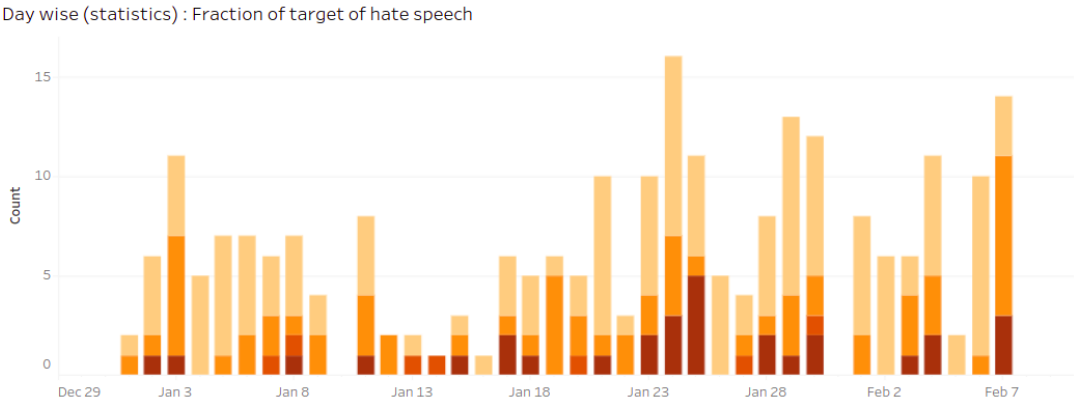
Regarding the category of harm, they have mainly identified the hateful content as either a direct or indirect political attack. Direct attacks specifically mention a political party or a politician. Whereas indirect attacks in the forms of taunts have no party or person mentioned but contain contextual evidence of political hatred. Additionally, not far from what we see in real-life electoral campaigns in India, some of the hateful content is not related to politics but are religious attacks. Given India's long history of politics and religion, the team added religious attacks as an additional category of harm. They also provide a breakdown of who are people propagating hate among political leaders (current cms and cm candidates for the electing states), political affiliates (members of the party who are not contesting but actively campaigning for their party), and common citizens (users who are not listed as apathy member but are strong proponents of the parties). The portal provides aggregated and daily statistics around these categories and targets of harm.

Based on the data curated so far, the team found that Delhi and Uttrakhand reported the second-highest number of hateful content (as geotagged by the tweets), right below Uttar Pradesh. This is an interesting observation as Delhi is the nodal center for 3 of our national parties (BJP, INC, and AAP). Despite Delhi not conducting elections, it still produces a significant volume of politically hateful tweets. Another observation made by the team is that direct attacks in political parties prevail over attacks on individual politicians. This can be attributed to calling out the party's name will creating more engagement from the opposing party and ordinary citizens. Around January end, there was an increase in the hateful content analyzed by the team. This attribute to the political rally conducted in Punjab around the same time. The team expects that as the frequency of physically coordinated rallies increases, the influx of political hate on Twitter will increase. They hope the ongoing data curation for February will reveal those insights. Meanwhile, the portal's aggregated and daily statistics are regularly updated.

Portal Link:
https://robinwatch.github.io/

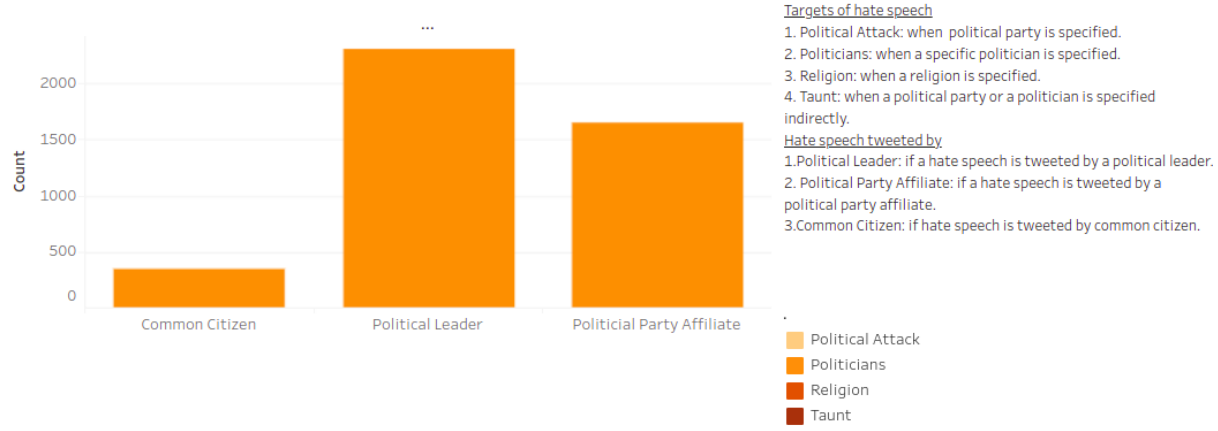
Some screenshots of the portal:



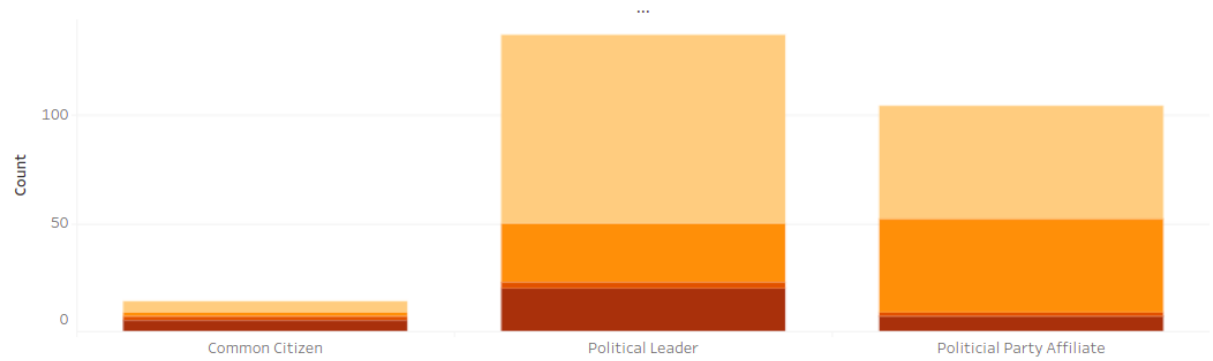


An overall and daily portion of the targets of political hate.

Aggregation of target of hate speech per user type



User type wise (statistics) : Fraction of target of hate speech



Breakdown of the instigators of hate against targets of political hatred.