

## **IDEA 1:**

### **Project title 1 (Descriptive title that captures your idea1)**

*Is there a difference in how user's responses to hateful content vs fake news?*

#### **What you want to do?**

Using hateful and misinformation on Twitter and Reddit (r/news), I want to understand if there is a significant difference in user activities (like share, upvote) and type of reply they post when they see hateful and misinformations.

#### **Method:**

- Data
  - o Verified hateful and misinformation, user activities
  - o Users comment and reply to hateful and misinformation posts.
  - o Use user activities on an accurate and non-offensive post as base-line
- Statistical tests on user activities
- Sentiment analysis of user replies/comments

#### **Why should we care?**

With the rise of social media and online communities, most people are reading news online as well as part of their social life is also online. Online platform providers need to detect hateful or fake information being shared on their platform. I believe this research would provide an insight into user behavior and see whether online users can detect hateful and misinformation that comes to their social media feed. If there is no difference, then Twitter needs to do more than just launching a service like [birdwatch](#). They need to make sure online users can effectively detect hateful content to make birdwatch a successful tool to protect communities.

**Keywords** – To mark the topic and domain of the idea.

*Fake news, hate speech, sentiment analysis, Reddit, Twitter*

**IDEA 2:****Project title 2 (Descriptive title that captures your idea2)**

*Automated detection of hateful content on Twitter using a context-aware model*

**What you want to do?**

*There is a lot of research done in this area using hateful and offensive lexicons. We know some words are more offensive than others. Hence just having a count of words makes it difficult to accurately detect whether text content is hateful or offensive. I want to improve hateful and offensive text classifiers by adding text contextual features. Similar research is done by [Lei Gao](#) using fox-news data.*

**Why should we care?**

*Accurate classification of hateful content in online communities is a big challenge. Bag of word and text style-based models fail to differentiate between offensive and hateful content. To control the spread of hateful content in online media, it is important to detect it accurately. I believe providing the complete context of threads, user, and part of speech tagged sentences would provide an additional feature that can be used to detect hateful content accurately. Having a model that can detect hateful content with short text like 140-200 characters tweets would help in controlling hateful and offensive content being shared on Twitter or other online communities.*

**Keywords** – To mark the topic and domain of the idea.

*Hate speech, offensive language, Twitter*