Data Annotation Guidelines

Steps:

1: Unzip the zip folders of the Collected memes.
2: Open *extracted_texts_pytesseract.xlsx* on drive
2: Fill up the *Actual Text* column with the text on the meme which is in the zip folders.
3: Analyze the meme and decide if it is hateful. If yes, select Hateful or else Not Hateful
4. **Bold the words that you think makes the meme hateful**

Characteristics of Hateful memes:

What could be considered hateful memes in terms of text:
1. Direct Insults or Name-Calling

- Content that includes derogatory or offensive terms aimed at an individual or group.
- Example: "Idiots like them should stay out."

2. Stereotyping or Generalizations

- Content that perpetuates harmful stereotypes or makes sweeping generalizations about a group.
- Example: "All men are lazy."

3. Incitement to Violence or Harm

- Content that encourages, glorifies, or justifies violence or harm against an individual or group.
- Example: "We should get rid of all the leftists."

4. Dehumanizing Language

- Content that describes individuals or groups as less than human, using animal comparisons or other demeaning metaphors.
- Example: "LGBTs are rats that need to be exterminated."

5. Discrimination or Exclusion

- Content that promotes or endorses unequal treatment, exclusion, or discrimination based on race, religion, gender, sexual orientation, etc.
- Example: "Only white deserve to be here."

6. Mockery or Ridicule

- Content that mocks, ridicules, or belittles individuals or groups, particularly in a way that is intended to demean or humiliate.
- Example: "Indian attitude vs indian reality"

7. Blaming or Scapegoating

- Content that unfairly blames a group for societal problems, historical events, or other issues.
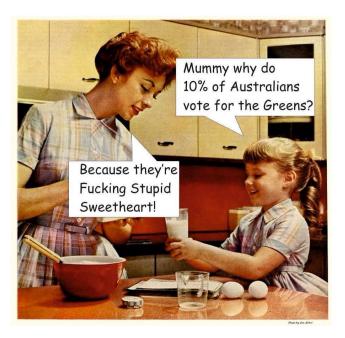- Example: "Palestinians are the reason for all our problems."

8. False or Misleading Information

- Content that spreads false or misleading information with the intent to harm a group's reputation.
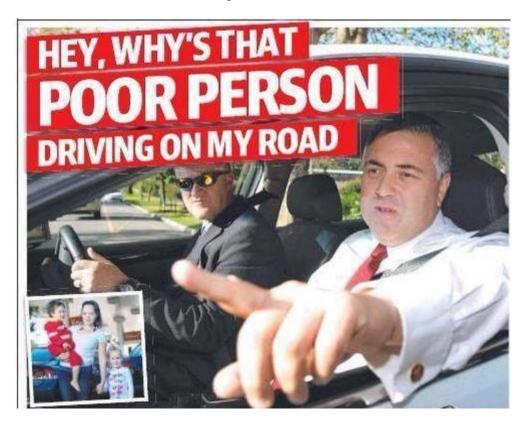- Example: "Immigrants steal jobs."

What could be considered hateful memes in terms of Image:

For analyzing the memes, both the text as well as the image are examined thoroughly. It is possible that sometimes the text is neural but the image is hateful or vice versa. Even in certain cases, none of the text nor the image is hateful but the meme represents a hateful meaning. Therefore, no emphasis is given to a single modality. The full meme is examined thoroughly. If the full meme resembles a hateful message, only then it is labeled hateful.

Since memes have both a textual component and an image component, it is mostly difficult to rely on a single modality to make a decision regarding the class label of the dataset. For instance, the image below shows when the image does not contain anything hateful. However, the text spreads hatefulness.

Similarly, text does not represent hateful content but the image does. In certain cases, neither the text nor the image resembles hateful meaning. However, when they are combined together, they

indicate a hateful message. The meme below is an example of it.

Target

1. **Targeted Individual (TI):** The hate directed towards a specific person (male or female) based on his fame, gender, race, or status. The person might be an artist, an actor, or a wellknown politician such as Huge Jackman, Pat Cummins etc.

2. **Targeted Organization (TO):** When the hate propagates towards any particular organization which is a group of people having certain goals such as a business company (e.g., Optus), government institution(e.g., School & Colleges), political organization (e.g., Federal), etc.

3**. Targeted Community (TC):** Hate on any specific group of people who hold common beliefs or ideology towards any religion (e.g., who follow the ideology of Buddhism), culture (e.g., who celebrates the valentines day), person (e.g., followers of cricketer Shane Warne) or organization (e.g., followers of Kmarts).

4. **Targeted Society (TS):** When a meme promotes hate towards a group of people based on their geographical areas such as mocking entire Indian people or British people it becomes hateful to an entire society.