Offensive Text Classification

|  |  |
| --- | --- |
| [1]Hate speech, harassment, prejudice, and discrimination | Text which is :-   * Hurting religious / racial sentiments (Name of a religion / race / religious person with negative sentiment) E.g. All Muslims are terrorists * Inciting / VIolence of threat to kill on the basis of a religion / race / ethnicity E.g. jews must be killed * Supporting or asking to join banned terror organisations or hate groups E.g Join ISIS * Attacking an individuals character E.g. Trump / Biden are \*\*\*\*   Hate directed at individuals and/or groups including:  • Attacks/ Insults/ slurs/ profanity  • Shaming or degrading content  • Criticism and mockery using insults and/or profanity  • Negative caricatures and depictions  • Hateful jokes  • Trivializing/ glorifying/ inciting violence  • Promoting discriminatory policies  Glorifying/supporting hate groups through banners and slogans  Use UN Designated Hate Groups |
| [2]Self-harm | Text Explaining how to or encouraging people to commit suicide or injuure themselves. E.g. 10 ways to kill yourself |
| [3]Extreme Violence | Text Explaining how to or encouraging people to commit acts of violence or threats to violence E.g. 10 ways to kill your enemy |
| [6]Profanity and Vulgar Language | Swear words, curse Eg: bitch get up off me; bullshit express; fuck, slut, etc.. |