

3. Automatic Detection of Online Nepali Abusive Text for Intimate Partner Violence Research

PROBLEM:

- Mostly we only look into physical abuse as a form of violence and As most of the young society is moving towards social engagements virtually via various social media and messaging platforms, there has been a rise in incidences of online abuse and violence. Violence against an intimate partner via internet is online Intimate Partner Violence (IPV), which seems to be increasing but its nature and prevalence is not yet well known.
- This work focuses on aiding IPV research to understand the nature and prevalence of online IPV, and to build foundations for detecting potential IPV Nepali texts at scale.
- One of the first steps to detect potential IPV Nepali texts is to be able to automatically detect abusive texts and contexts in Nepali language. However, there currently exists a gap in good Nepali Natural Language Processing (NLP) AI models

OUR PROPOSAL:

- Build a suitable annotated dataset and train AI models for abusive text detection. The texts are gathered from Twitter posts and YouTube comments, which are annotated to describe various forms of abuse. Train the AI model to generate a rough estimate of how prevalent is Intimate Partner Violence in the context of Nepal. comments, which are annotated to describe various forms of abuse.



PAPER PUBLICATION DETAILS:
In Progress

- Train the AI model to generate a rough estimate of how prevalent is Intimate Partner Violence in the context of Nepal.
- An online web-based platform that collects publicly available data and provides outputs of model for humans to verify and correct, when necessary, from which various studies can be carried out to understand the nature and prevalence of abusive texts.

PROJECT STATUS:

- A chat application has been developed where users can voluntarily simulate conversations, which is being used to simulate IPV related and normal conversations.
- Annotated Nepali Twitter posts at sentence and phrase level for various forms of abusive texts vs normal texts
- Evaluation of AI models is ongoing.
- A web-based platform has been built which gathers keywords based Nepali texts from Twitter and youtube comments, classifies them using AI model into abusive vs non-abusive, detects phrases and classifies into various types of abuse, human in the loop system to verify and correct AI predictions, and analytics on various forms of abusive text.

FUNDED BY:

Sexual Violence Research Initiative (SVRI)



PROJECT TEAM:

Rabin Adhikari & Dr. Bishesh Khanal,
members from ChildSafeNet



CLINICAL PARTNERS:

- ChildSafeNet Nepal

ChildSafeNet

Safer internet for children and young people