1 Hinduphobic COVID-19 X (Twitter) Dataset

Our **Hinduphobic COVID-19 X (Twitter) Dataset** is a meticulously curated collection of tweets sourced from X (formerly Twitter) between April 2020 and January 2024, specifically focusing on sentiments related to Hinduphobia during the COVID-19 pandemic. This dataset is uniquely ours, developed to capture a comprehensive view of Hinduphobic discourse. It predominantly includes tweets from India, written in Hinglish (a mix of Hindi and English) and English. The dataset consists of 1,000 manually labeled negative tweets (Hinduphobic) and 1,000 positive/neutral tweets (Pro-Hindu), offering a balanced foundation for sentiment analysis. Tweets were collected using the X (Twitter) API, employing targeted keywords and hashtags to identify content explicitly linked to the Hindu community during the pandemic. Rigorous preprocessing steps were applied to enhance the dataset's quality and relevance, including the removal of irrelevant content, spam, duplicate entries, web links, hashtags, and profile tags. Contractions were expanded, emojis were converted into text, and non-text symbols were transformed for consistency. A **Human-in-the-Loop (HITL)** approach was utilized for manual labeling, ensuring the accuracy of sentiment classification. The final dataset comprises 8,000 tweets, with 4,600 labeled as Hinduphobic and 3,400 as positive/neutral. This dataset plays a pivotal role in training and fine-tuning advanced models like **HP-BERT**, a specialized BERT model designed to enhance the detection of Hinduphobic content and address nuanced expressions of hate speech.

For further details and to access the dataset, visit our Kaggle repository: **Hinduphobic COVID-19 X (Twitter) Dataset**[1].

1.1 Keywords for Scraping the Hinduphobia Dataset from X (Twitter)

The keywords were selected to capture a comprehensive range of sentiments (negative, positive, and neutral) associated with Hindu beliefs, practices, and community efforts during the pandemic.

1.1.1 Negative Keywords

These terms are commonly used in a derogatory context to mock or stereotype Hinduism, often perpetuating negative perceptions about Hindu culture and practices:

- **Brahmanism** – Critiques Hindu social structures, stereotyping Hinduism as caste-based.

- **Cow piss drinker** – Mocks the reverence of cows, implying backwardness.

- **Cow worshipper** – Belittles Hindu practices involving cows, an important religious symbol.

- **Devil worshipper** – Falsely associates Hinduism with demonic practices.

- **Dothead** – Racial slur mocking traditional bindi-wearing Hindus.

- **Hindu fatalism** – Misrepresents Hindu beliefs as passive acceptance of suffering.

- **Dung worshipper** – Ridicules cultural uses of cow dung, portraying it as primitive.

- **Hindutvavadi** – Used disparagingly for supporters of Hindu values, implying extremism.

- **Bhakt** – Labeling Hindu devotees as fanatical.

- **Idol worshipper** – Criticizes image veneration, portraying it as superstitious.

- **Kaffir** – Pejorative term for non-believers, used derogatorily against Hindus.

- **Cow urine cure** – Implies Hindus rely on unscientific treatments.

- **Cow dung cure** – Mocks traditional medicine practices involving cow dung.

- **Kumbh Mela** – Targeted as a "super-spreader" event, disproportionately criticized.

- **Temple gatherings** – Described as reckless during the pandemic.

- **Sacred cow** – Mocks the reverence for cows, implying irrationality.

---

[1] https://www.kaggle.com/datasets/ashutoshsingh22102/hinduphobic-covid-19-x-twitter-dataset-india

- **Superstitions** – Dismisses Hindu rituals as backward or unscientific.

- **Backward beliefs** – Stereotypes Hindu customs as primitive.

- **Hindu rituals** – Portrays Hindu practices as irrational.

- **Temple practices** – Derides temple worship as out of touch.

- **Snake worship** – Ridicules animal worship in Hindu culture.

- **Primitive religion** – Dismisses Hinduism as outdated.

- **Outdated practices** – Labels Hindu rituals as irrelevant.

- **Mythology-based religion** – Implies Hinduism lacks factual basis.

- **False gods** – Disrespects Hindu deities.

- **Voodoo religion** – Stereotypes Hinduism as magical or superstitious.

- **Fake religion** – Delegitimizes Hinduism.

- **Barbaric rituals** – Depicts Hindu rituals as savage.

- **Stagnant society** – Frames Hindu society as resistant to change.

- **Backward people** – Labels Hindus as regressive.

1.1.2 Positive Keywords

These terms reflect the positive impact of Hindu community efforts, charity, and unity during the COVID-19 pandemic, showcasing Hindu values of compassion, service, and resilience:

- **NamasteForSafety** – Encourages the Hindu greeting as a contact-free gesture.

- **HindusForHumanity** – Highlights Hindu humanitarian efforts.

- **ISKCONRelief** – COVID-19 relief efforts by ISKCON, emphasizing community support.

- **RSSHelpingHands** – Reflects contributions by the RSS for essential services.

- **TempleAidCOVID** – Temples providing food, medicine, and shelter.

- **TempleDonations** – Temple-led charity drives for COVID-19 relief.

- **HinduCOVIDRelief** – Organized relief efforts by Hindu organizations.

- **HinduCharity** – Widespread charitable contributions by Hindu communities.

- **PlasmaDonation** – Hindu involvement in plasma donation drives.

- **CommunityService** – Promotes selfless service to local communities.

- **ServiceToHumanity** – Hindu values in serving humanity.

- **SwaminarayanRelief** – Relief efforts by the Swaminarayan Hindu community.

- **MahavirTempleAid** – Charity by Mahavir Temple.

- **HinduUnity** – Emphasizes collective support within Hindu communities.

- **HinduSolidarity** – Unity and network-building for aid.

- **COVIDReliefEfforts** – Coordinated Hindu efforts for pandemic assistance.

- **PandemicRelief** – Hindu contributions to distributing essentials.

- **HealingThroughFaith** – Role of Hindu spirituality for emotional resilience.

- **ReliefForAll** – Inclusive Hindu aid efforts.

- **TogetherWeServe** – Joint efforts to serve during the pandemic.

- **HinduCommunityHelp** – Local Hindu efforts to organize resources.

- **BhaktiForHumanity** – Extending Hindu devotion to serve humanity.

- **SpiritualStrength** – Hindu belief in resilience.

- **SewaInternational** – Global Hindu charity providing COVID-19 relief.

- **SanatanSewa** – Service efforts based on Sanatan Dharma.

- **FoodDistributionDrive** – Community efforts in food distribution.

- **MaskDonation** – Hindu donation drives for protective equipment.

- **GharGharAnaj** – Hindu-led food and essentials distribution.

- **AnnapoornaScheme** – Food drives inspired by Hindu values of nourishment.

- **HelpingHandsHindu** – Community readiness to serve during crises.