

Jhatta Samachar Dataset Requirement

Dataset Format

The dataset should consist of news articles scraped from various sources and be provided in either JSON or CSV format. Each news article should have the following fields:

- **id**: A unique identifier for the news article, ideally the URL from which the news is extracted. This allows for easy future retrieval of additional information.
- **title**: The headline of the news article.
- **content**: The full text of the news article.
- **summary**: A concise summary of the article's content. The summary should accurately reflect all key information from the article to ensure quality outputs.
- **category**: The category of the news, which can typically be found in the article's URL (e.g., <https://ekantipur.com/en/news/some/news/address/here/>).
- **author** (optional): The name of the source of the news. This can be useful for attribution or filtering news by a specific source in the future.
- **publish_date** (optional): The date the article was published. This can be important for time-sensitive applications or analyses. This can be obtained from the url of the news as in <https://ekantipur.com/en/news/2024/05/06/some/news/address/here/>
- **language** (optional): The language in which the news article is written. This could be helpful if we plan to work with multilingual content.

Example Format

In JSON

Here's an example of how the dataset could be structured in JSON:

```
[  
  {  
    "id": "https://ekantipur.com/en/news/2024/10/06/aid-to-orthopedic-hospitals-by-chemicals-and-medical-supplies-association-03-07.html",  
    "title": "Aid to Orthopedic Hospitals by Chemicals and Medical Supplies Association",
```

```

    "content": "The Chemical and Medical Supply Association has supported the Nepal Orthopedic Hospital in Jorpati. \n\nBhakta Bahadur Hamal, president of the association, said that on the occasion of its 20th anniversary, the association has supported essential chairs for sick people as part of its social responsibility.",
    "summary": "The Chemicals and Medical Supplies Association donated essential chairs to Nepal Orthopedic Hospital as part of their 20th anniversary celebration and social responsibility.",
    "category": "news",
    "author": "Ekantipur",
    "publish_date": "2024-10-06",
    "language": "en"
  }
]

```

In CSV

Here's the same example in CSV format:

```

id,title,content,summary,category,author,publish_date,language
https://ekantipur.com/en/news/2024/10/06/aid-to-orthopedic-hospitals-by-chemicals-and-medical-supplies-association-03-07.html,Aid to Orthopedic Hospitals by Chemicals and Medical Supplies Association,"The Chemical and Medical Supply Association has supported the Nepal Orthopedic Hospital in Jorpati.\n\nBhakta Bahadur Hamal, president of the association, said that on the occasion of its 20th anniversary, the association has supported essential chairs for sick people as part of its social responsibility.",
"The Chemicals and Medical Supplies Association donated essential chairs to Nepal Orthopedic Hospital as part of their 20th anniversary celebration and social responsibility.",news,Ekantipur,2024-10-06,en

```