<https://www.kaggle.com/datasets/maxscheijen/dutch-news-articles>

**About Dataset:**

Dutch News Articles

This dataset contains all the articles published by the NOS as of the 1st of January 2010. The data is obtained by scraping the NOS website. The NOS is one of the biggest (online) news organizations in the Netherlands.

**Features:**

datetime: date and time of publication of the article.

title: the title of the news article.

content: the content of the news article.

category: the category under which the NOS filed the article.

url: link to the original article.

<https://www.kaggle.com/datasets/harrotuin/dutch-news-headlines>

**Context**

This dataset contains news headlines of two Dutch news websites. All sarcastic headlines were collected from the Speld.nl (the Dutch equivalent of The Onion) whereas all "normal" headlines were collected from the news website Nu.nl.

**Content**

Next to the news headlines the dataset contains a sarcasm label, the source, and the news subject. News on three different subjects was collected: politics, foreign, and domestic news. Some headlines appear multiple times because they are listed under multiple subjects. For some applications, it might be useful to delete duplicates. The following columns are available:

Headline: the original news headline

link: the URL to the original news article

source: either speld.nl or nu.nl

is\_sarcastic: boolean (speld being sarcastic and nu.nl not)

is\_binnenland: boolean for news category which translates to "domestic news"

is\_buitenland: boolean for news category which translates to "foreign news"

is\_politiek: boolean for news category which translates to "political news"

<https://webz.io/free-datasets/dutch-news-articles/>

Get free Dutch news articles dataset crawled from the Webz.io API Language category.

Crawled Date Oct, 2015

Format Available JSON

<https://www.delpher.nl/nl/kranten> (probably not usable because of the old age of the news)

The collection comprises almost 2 million newspapers dating from 1618 until 1995. They are editions from every year of publication of the main Dutch national newspapers, such as De Telegraaf, De Volkskrant and Het Parool, supplemented with a selection of regional and colonial newspapers. The dataset consists of scans of the printed pages, with OCR and word coordinates.