**Project Name:** News Aggregator

**Github Link:** https://github.com/projectsforstudents2022/News-Aggregator.git

**Why was this project created?**

A news aggregator is essentially a piece of web software that compiles breaking news and events from many sources into one convenient location. A news aggregator is highly helpful in saving time consumption because it gathers all of the news that would otherwise need to be studied on several websites into one spot. Additionally, condensing this content will undoubtedly save the readers' time.

**What problem is it solving?**

The primary objective of this project is to create a news aggregator that can compile articles that are pertinent to a given input keyword or keyphrase. After improving the text, summarize the pertinent articles to provide a clear and concise summary for the reader. A news aggregator organizes and disseminates information from several sources.

**Entire explanation of project**

* **PROPOSED APPROACH**

Assembling relevant articles based on the user's entered keyword or phrase. The pieces are compiled from a variety of reliable sources, including the BBC and the World Health Organization. The pre-processing stage entails performing particular operations on the collected articles, including Lowercase is employed to limit the quantity of the vocabulary in our data that results in many instances of the same word meaning, Stop-word Removal is carried out to eliminate unimportant textual details in order to concentrate on keywords, and Lemmatization and Stemming are other techniques. Take the word out of its inflection and transfer it to its fundamental form.

Following the previous step of creating the similarity matrix, we turn it into a graph with edges that are defined by how similar the two objects are to one another. To determine the weight of the vertices, those edges are employed. The number of edges that are represented as a score for each vertex determines how important a sentence is. The user will be able to browse news from various sources thanks to the suggested system overview. The key feature of the system, a comprehensible summary of all the collected content on the same subject, will also be provided to the user.

Algorithm for model :

**Step 1:** Dataset is imported

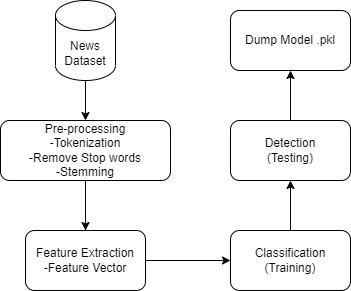
**Step 2:** The data is preprocessed, and the dataset is split into training and testing.

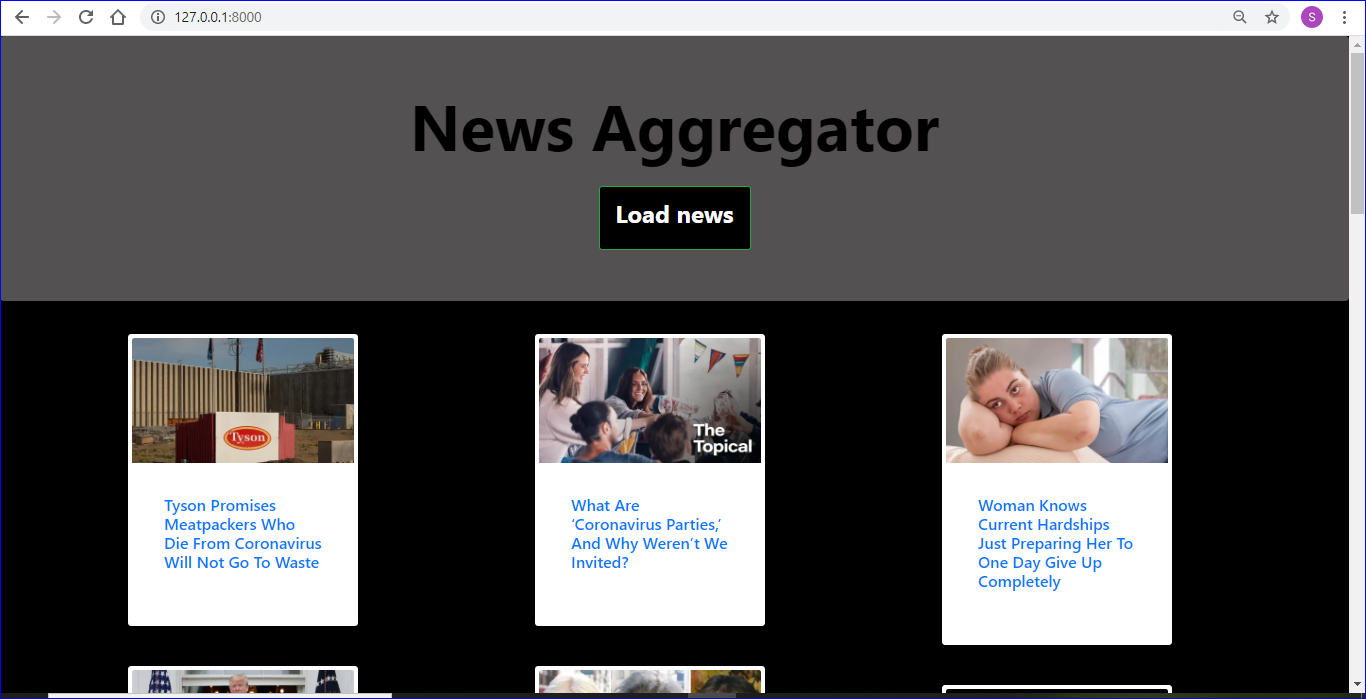
**Step 3:** The training data is converted to numerical using the vectorization approach.

**Step 4:** The training dataset is used to develop the prediction model, which includes the Passive Aggressive Algorithm, Random Forest, and other classification methods.

**Step 7:** A confusion matrix is created.

* **DATA FLOW DIAGRAM**



* **RESULT**
* **CONCLUSION**

Following the comparison of two summarization methods, the TextRank algorithm was selected above the Word Frequency strategy. The TextRank algorithm was used for this project because it provides readers with more effective summaries. The system will produce an output summary of information from online sources that includes the important concepts related to a specific article topic and can be comprehended without consulting the original article.