International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, March-April-2021, ISSN (Online): 2395-566X

News Classification System Based on Area

Prof. Krishnanjali Shinde, Prachi Prajapati, Sakshi Tagalpallewar, Ruchita Bacchuwar
IT Department
Genba Sopanroa Moze College,
Balewadi Pune
prof.krishnanjali@gmail.com, prachiprajapati830@gmail.com, sakshitagalpallewar@gmail.com,
ruchitabacchuwar1234@gmail.com

***Abstract*- In 21st century the internet is filled with loads of news articles, there is a pressing need to classify news according to the requirements of an individual. People are generally more interested what is going on, in their immediate surroundings. News has a vital role in the society. Most people read news every day to keep up with the latest information and trends. The information could be anything, from technology, disaster, politics, even the affair of the celebrities. After they absorb the information and understand it, it will be used by the people as a reference to their ideology and decision making. With the help of technology advancements, news disseminates relatively quick across the globe. Using the internet, people can send information from another side of the world in under a second. Because of this, almost any kind of information such as knowledge, idea, entertainment, and news from the people can easily spread to the community. With the development of the web, and ton of internet sites that provide similar information and data. So, users often discover it hard todecide that of those websites will offer thespecified information inside the foremost valuable and effective way.**

***Keywords*- Machine learning, RSS feeds, web scrapping, NLP, chatbot.**

I.INTRODUCTION

In the previous few years, the globe had unimaginable and huge growth within the rate of reports that's published. Individuals sleep in a time jam-packed with info, data, and news. So, today news has a vitalpart and position inside the community. As peopleread the news daily to stay up with the foremost recent data and inputs. This information is also regarding technology, sports, weather, food, and celebritiesor many different fields simply think about all thevarious websites that are out there and also the info they provide come in.

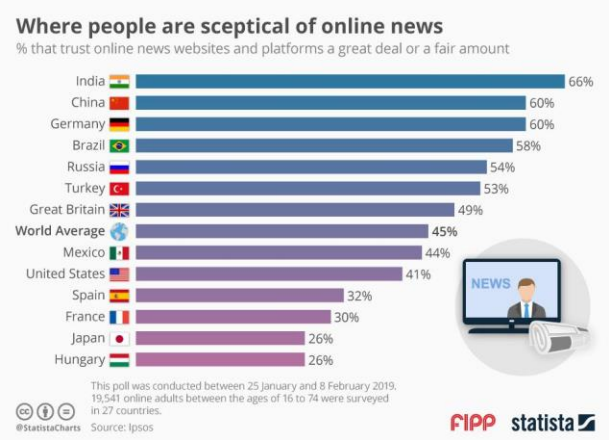

Fig 1. Graph 2

Visiting each single one that you're interested in will be boring and long task. Most of the people don't have enough time to try and do it, but there are solutions to it. That's where news aggregators.

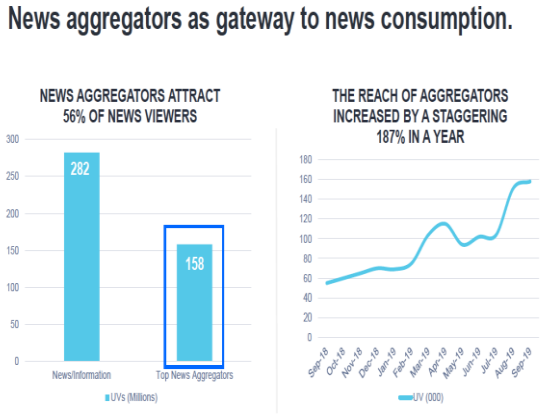

Fig 2. Graph 2

News aggregator is just online software which collects new stories and events round the world from numerous sources tied one place. News aggregator plays a really necessary role in reducing time consumption, as all of the news that might be explored through quite one web site is placed solely in a single location. Also, summarizing this aggregate content fully can save reader's time.In this paper; we tried to solve this issue by classifying the news articles on basis of cities and providing the entity with the gathering of town specific news.

We've developed our own net crawler forcontent extraction from the hypertext mark-up language pages of reports articles. Random Forests, Naive Thomas Bayes

© 2021 IJSRET

and SVMclassifiers are used for classification of reports articles. Results exhibit that machine learning techniques will be controlled to realize our goal and thus demand more analysis to enhance the efficiency of resolution this issue.News aggregator websites enable users to look at news and updates from various sources at one convenient location. They fetch the info, organize them in tags / classes, and show it within the right order for easier consumptions. Some well-liked news aggregation websites are FEEDLY, Google news, ALLTOP, News 360, Panda, TECHMEME, Flipboard, Pocket, etc.

The Best News Aggregators

NAME	FOUNDED	TOPICS	CONTRY	PLATFORMS
Google News	2002	Current events	The USA	Android, iOS, Web
Feedly	2008	Tech, marketing, design, fashion, food, etc.	San Francisco, California	Android, iOS, Web iOS, Android, Windows, Windows Phone
Flipboard	2010	Social-network aggregation	Palo Alto, California	
Pocket	2007	Online bookmarking, saving articles for later reading	San Francisco, California	Android, BlackBerry, iOS, Kindle Fire, Kobo, macOS, web browsers, Windows, and Windows Phones
City FALCON	2014	Investing, Trading, Technology, and Stock markets	The UK, London	Android, iOS, Web

Fig 3. **Text Here Fig Title.**

II. RELATED WORK

Existing systems provides news classification based on areas but not to certain extent they lack in many ways. Most of famous news aggregator system focuses on classifying news articles on basis of genres like entertainment, fashion, sports, politics, food, health, technology, etc.

Very few websites classifies them on basis of areas and as well as provides the chatbot application for better user experience. Random Forest Classifier, Support Vector Machine and Multinomial Naive Bayes are the machine learning techniques used by existing news classification system for classification of news.

III. PROPOSED SYSTEM ARCHITECTURE

1. Web Scrapping:
Web scrapping also called as web harvesting or web data extraction is process of collecting or extracting information or data from different websites. It will directly access the World Wide Web using hyper text transfer protocol or web browser. This technique can be implemented manually as well as automated process .For automated process into make use of bot or web crawler. Web crawler is additionally referred to as spider bot it's an internet bot that consistently browses the World Wide Web usually for net indexing purpose.

2. RSS Feeds:
RSS stands for Really simple Syndication andit's may be an easy, standardized content distribution method that may assist you keep up-to-date together with yourfavorite news casts, blogs, websites, and social media channels. Rather than visiting sites to search out new posts or subscribing to sites to receive notification of latest posts, one will scan RSS posts through RSS post column in news websites and then reading in RSS reader.

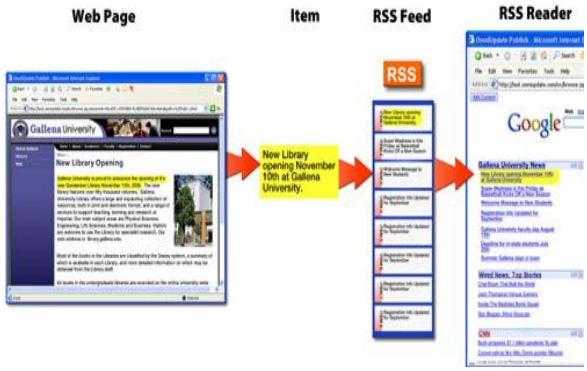

Fig 4. **Text Here Fig Title.**

3. Random Forest Classifier:
Random Forest isasupervised classification algorithmic rule that was 1st introduced properly by Leo Brieman. It uses varietyof uncorrelated decision tree classifiers and fits them on various sub samples of dataset. Every sub sample dataset for decision tree has same size as the original dataset. Random forest Classifiers uses boot strap aggregating techniques that repeatedly selects random sample of the coaching set with replacement and uses this sample to create trees learn since boot strap technique decreases the variance and therefore leads to higher performance. Random Forest classifier is employed in outlier detection and replacing missing knowledge. It's scalable because it will run on giant data sets.

4. Naive Bayes:
Naive Bayes classifier is a popular method for text classification problems where the classifier has to classify the articles given in the given document into different category or columns. It was first proposed by D. Lewis. Naive Bayes classifier is based on probabilistic technique of classification which derives its roots from Bayes Theorem. This technique is based on the assumption that the independence exits between the various features. It is a scalable classifier and can run efficiently with large data sets. Naive Bayes classifier is fast as compared to other

© 2021 IJSRET

classifiers and thus is used as a baseline for text classification problems.

5. SVM Classifier:
SVM (Support Vector Machine) works on the principle of Supervised Learning. SVM needsa training set and labels related to it. Aftertraining, if a test data is fed in, the model assigns it toone class or the other class. It performs well with linear classification. It can even work efficiently on a nonlinear classification using a kernel trick by mapping the inputs into high dimensional feature space. It constructs a hyperplane for classification the hyper-plane is chosen such that the spacebetween the closest data point on either side ismaximized.

6. Chatbot:
A chatbot is a software application used to conduct chat communications via text or text-to-speech, in context of providing direct contact with a live human agent. Chatbot was designed to convincingly simulate the way a human would behave as a conversational partner, chatbot systems typically require continuous tuning and testing.A chatbot is anArtificial Intelligence (AI) packagethatcan simulate a spoken communication with a user in naturallanguage.

It's an advanced implementation of natural language processing, taking us closer to communicating with computers in a way similar to human-to-human conversations. Chatbots are extremely helpful for users for users which are interacting with totally new system. Chatbot provides much needed assistance to user. It helps user to easily navigate through system and also makes user aware of unknown functionalities of system.

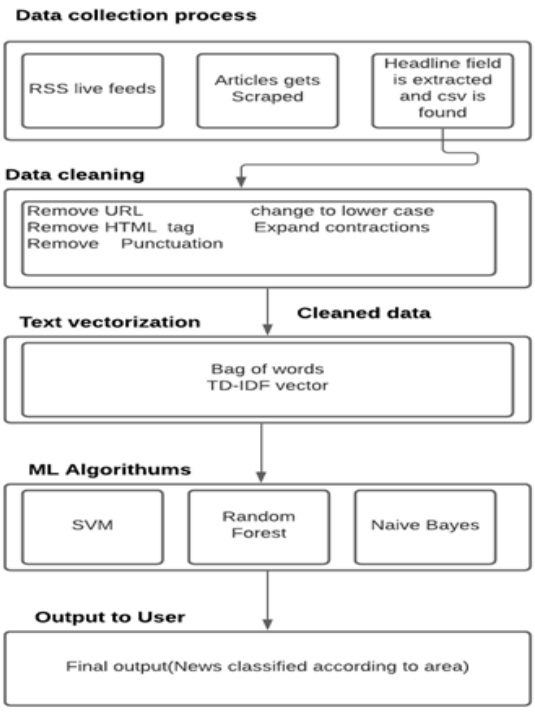

Fig 5. Proposed System Architecture.

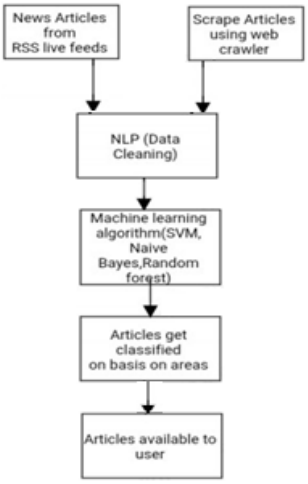

Fig 6. Flowchart of system.

7. Advantages of Proposed System:

- Proposed system will classify news articles which are scraped using web crawlers on basis of areas like city, state, and country using machine learning algorithms.
- System will also classify news according to genres like sports, business, politics, entertainment, fashion, etc.
- News articles are scraped from RSS live feeds of 100+ news websites
- It provides a chatbot application for easy communication and for reciting top 10 news for users anytime anywhere.
- Users can register to publish their blogs as well and make use of other features of website.
- Users can rate articles as per their convenience.
- This system also provides local newspapers to read.
- Easy navigation and simple to use website.

IV. CONCLUSION

In this survey paper we have specified the design and system structure of news aggregator system that classifies news on basis of area (locations). It provides news based on particular area like city, state, and country.

The proposed system also comprises of a chatbot that helps users to solve their queries in no time and provides best user experience. Users can view their search history, write reviews on articles, like or dislike articles, chat with other users having similar interest areas and can also publish articles on system.

Live chat window for users having same area of interest can be implemented as future scope.

REFERENCES

[1] Vighnesh Rao, Jayant Sachdev, “A machine learning approach to classify news articles based on location”.

© 2021 IJSRET

IEEE Xplore Compliant - Part Number: CFP17M19-ART, ISBN: 978-1-5386-1959-9.

[2] Zaiyinh Wang, Bahoa Song, “Research on hot news classification algorithm based on deep learning”.2019 IEEE 3rd Information technology, Networking, Electronics and Automated Control Conference.

[3] D. Shen, Z. Chen, Q. Yang, 3.H. Zeng, B. Zhang, Y. Lu, and W. Ma,Web-page classification through summarization, in Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2004, pp. 242249

[4] Bahana, R. Adinugroho, R. Gaol, F. L. Trisetyarso, A. Abbas, B. S. & Suparta W. (2017). Web crawler and back-end for news aggregator system (Noox project). 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom). doi:10.1109/cyberneticscom.2017.8311684

[5] S. Chowdhury and M. Landoni, "News aggregator services: user expectations and experience", Online Information Review, 30(2), pp.100-115, 2006

[6] Scrapinghub, Ltd., "Scrapy," GitHub repository, 6 December 2016. [Online]. Available: https://github. Com/scrapy/scrapy. [Accessed 27 March 2017].

[7] L. Richardson, "Beautiful Soup Documentation - Beautiful Soup 4.4.0 documentation," Crummy, 2 February 2012. [Online]. Available: https://www. crummy.com/software/BeautifulSoup/bs4/doc/. [Accessed 27 March 2017]. [11]

[8] E. Bower, "linkGrabber", GitHub repository, 10 July 2015. [Online]. Available: https://github.com/ Michigan-com/linkGrabber. [Accessed 2017 March 27]

[9] Manning, C. and Schutze, H., Foundation of statistical natural language processing Cambridge, Mass: MIT press, 1999.

[10] C. Ee and P. Lim, Automated online news classification with personalization.

[11] D. Lewis, “Naive (bayes) at forty: The independence assumption in information retrieval,” Machine Learning: ECML-98, pp. 415, 1998.

© 2021 IJSRET