

CSE3024- Web Mining

Digital Assignment - III

News Classification and Recommendation System

By

20BCE1184
20BCE1730

Santhoshkumar G
B Yeshwanth

B.Tech CSE

Submitted to

Dr.A.Bhuvaneswari,
Assistant Professor Senior,
SCOPE, VIT, Chennai

School of Computer Science and Engineering



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

November 2023



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

VIT Chennai

Vandalur - Kelambakkam Road, Chennai - 600 127

WINTER SEM 22-23

Worklet details

Programme	B.Tech with Specialization	
Course Name / Code	Web Mining / CSE3024	
Slot	C1+TC1	
Faculty Name	Dr.A.Bhuaneswari	
J Component Title	News classification and Recommendation System	
Team Members Name Reg. No	20BCE1184	Santhoshkumar G
	20BCE1730	B Yeshwanth

Team Members(s) Contributions – Tentatively planned for implementation:

<i>Worklet Tasks</i>	<i>Contributor's Names</i>
Database connection and integration using Pymongo	Yeshwanth
Preprocessing	Yeshwanth, Santhoshkumar
Model building	Yeshwanth
Visualization	Yeshwanth, Santhoshkumar
Technical Report writing	Yeshwanth, Santhoshkumar
Presentation preparation	Yeshwanth, Santhoshkumar

ABSTRACT

News is an important factor in everyone's daily life. The news we read sometimes makes our mood. But we need to classify news based on their categories. Not every news can be given to the user. So, we will develop a newsclassification model to classify the news into different categories. Similarly, we can provide news suggestions based on the similarity present between a given news and a list of news. In today's fast-paced digital landscape, news plays a pivotal role in shaping our understanding of the world around us. It not only informs us but also has the power to influence our emotions and opinions.

However, with the vast amount of news articles available online, it becomes imperative to streamline and categorize this information for users. This is where web mining, a powerful data extraction and analysis technique, comes into play.

Web mining enables us to develop a News Classification Model that categorizes news articles into distinct and relevant categories. By harnessing natural language processing and machine learning algorithms, this model can automatically classify news based on their content, whether it's politics, sports, technology, or entertainment. This categorization not only facilitates efficient content organization but also ensures that users receive news that aligns with their interests and preferences. Furthermore, web mining allows us to offer personalized news suggestions by measuring the similarity between a given news article and a curated list of news stories. This similarity analysis leverages text mining techniques to identify shared themes, keywords, or sentiments, enabling us to recommend additional articles that users are likely to find interesting and informative.

In essence, web mining empowers us to enhance the news consumption experience. It enables us to classify news articles accurately, ensuring that users receive content that resonates with their preferences. Additionally, by providing tailored news suggestions based on content similarity, we can keep users engaged and informed in a personalized and efficient manner. This synergy between web mining and news classification and recommendation systems brings us closer to a more efficient and satisfying news browsing experience in the digital age.

1. Introduction

In today's world, user experience and user preferences matter a lot. The ads we see, the recommendations we get in social media, the news we see, the feeds we get all are related to our previous searches. Our data is monitored and we get suggestions related to our previous data. We apply this to our project where we classify news into different categories based on their content and we suggest news. Our project uses machine learning algorithms to learn about the news with their categories. We develop a model based on the text present by taking out the words present in the description. For this we tokenize the words and there comes our Natural Language Processing part. Then we can suggest news by using similarity measures. Here we use Cosine similarity measure to identify the similarity between a given document and a list of documents. The similar documents can be identified and with this model, news can be suggested based on the users' recent searches. In today's digital landscape, user experience and personalization are paramount. From targeted advertisements to tailored social media recommendations, the online world revolves around catering to user preferences and behaviors. This paradigm extends to news consumption as well, where web mining and machine learning techniques are leveraged to classify news and offer personalized news suggestions. Our project is centered around enhancing the news consumption experience by classifying news articles into distinct categories based on their content. To achieve this, we employ web mining techniques to gather a diverse range of news articles from various online sources. This extensive dataset becomes the foundation for our News Classification Model, which employs machine learning algorithms. The core of our project relies on Natural Language Processing (NLP) techniques, where we tokenize the words in news articles, enabling the model to understand the underlying textual content. This NLP component plays a pivotal role in categorizing news into predefined topics

such as politics, sports, technology, and more. By analyzing the content, our model becomes adept at associating news articles with their respective categories.

Moreover, our system takes personalization a step further by employing similarity measures, such as the Cosine similarity measure. This measure allows us to identify the resemblance between a given news article and a list of previously read or recommended articles. By assessing similarities in themes, keywords, and sentiments, our system suggests additional news articles that align with the user's interests, effectively curating a personalized news feed. In essence, web mining and machine learning techniques converge in our project to revolutionize news consumption. By classifying news articles accurately and suggesting content based on user preferences, we are not only making news more accessible but also enhancing the overall user experience. Our model ensures that users receive news that resonates with their interests, keeping them engaged, informed, and satisfied in today's data-driven digital age. This integration of web mining and machine learning represents a significant step towards a more user-centric and efficient news browsing experience.

2. Literature Survey

Sl no	Title	Author /Year	About the Paper
1	Social Media Recommender Systems: Review and Open Research Issues	Anitha Anandhan; Liyana Shuib; Maizatul Akmar Ismail; Ghulam Mujtaba 27 February 2018	RSs are developed based on user textual reviews, ratings, and comparative opinions. RSs for social media resources, such as blogs, forums, social network websites, social bookmarking websites, video portals, and chat portals help users to collaborate effectively.

			<p>Social media resources are used in the RS for recommending contents, articles, news, ecommerce products, and users. his paper aims to provide a comprehensive review of the social media RS on research articles published from 2011 to 2015 by exploiting a methodological decision analysis in six aspects, including recommendation approaches, research domains, and data sets used in each domain, data mining techniques, recommendation type, and the use of performance measures.</p>
2	News recommendation system using machine learning	<p>Neha Rani, Sudhir Sudhir Pathak</p> <p>2018-06-05</p>	<p>The forecasting of financial news is yet becoming the main issue to divide the new into different classes on the basis of present time series. Moreover, it might be utilized for predicting and analyzing the stock market for the particular industry. Thus, the new content is significantly important to influence market forecast report. In this paper, the financial news from four countries namely America, Australia, India and South Africa along with their stop words</p>

			<p>are consider. The words along with their weighted values are determined and then the neural network is trained. Here, artificial neural network is used for classifying the appropriate results for the given input data. At last, the comparison of ANN with SVM is shown.</p> <p>Experiments show that the ANN classification provides high accuracy to predict the news than the SVM classifier.</p>
3	Adversarial Training for Fake News Classification	<p>Abdullah Tariq, Abid Mehmood, (Member, Ieee), Mourad Elhadeif, (Member, Ieee), Muhammad Usman Ghani Khan</p> <p>29 July 2022</p>	<p>Nowadays, there are a lot of fake news prevailing in social media platforms. Many people get a false view on many things due to the fake news present. So, this Journal covers an approach for fake news detection. This paper uses transformer encoders to get statistical information on some input text. Inputs are being perturbed in Adversarial Training so that model misclassifies the given inputs. Fast Gradient Sign Method (FGSM) is one of the methods proposed in to generate the perturbed examples. These perturbed samples are</p>

			<p>called adversarial samples. Input news text is passed to the transformer model that comprises an embedding layer and a series of hidden layers. The hidden state of [CLS] token representing the whole input sequence is then passed to the classification layer, and then the loss is computed. In order to estimate the perturbation, we compute the gradient of the loss in terms of the embedding matrix. The adversarial example is created by adding this perturbation to the transformer model's embedding matrix. Then, this adversarial example also goes through a series of hidden layers, then the classification layer, and finally, the loss is computed. The total loss becomes the sum of losses of clean and adversarial examples. Then the backward step is taken, and the model's parameters are updated. In this way, the fake news classification model is trained.</p>
--	--	--	---

4	Big Data ML-Based Fake News Detection Using Distributed Learning	Alaa Altheneyan; Aseel Alhadlaq 22 March 2023	<p>Fake news gives people many fake information. A system to detect fake news, if present will be very good. Social networking sites like Facebook and Twitter generate daily data [15] In 15, fake news detection using machine learning is done in 3 languages. So, this Journal covers a system which detects fake news. The dataset comprises occurrences with a title, article body, and one of the four labels “Disagree”, “Agree”, “Unrelated”, and “Discuss”. In pre-processing, labels are encoded into numeric target values and perform some pre-processing steps. Preprocessed data is split into 75% data for training and 25% for testing. The augmented corpus is created by combining 49972 stances with 1683 bodies based on ids. The corpus has four distinct classes (agree, disagree, discuss, unrelated). It contains 8909 discuss</p>
---	--	---	---

		<p>stances, 36545 unrelated stances, 3678 stances, and 840 disagree stances. After gathering headlines and articles in one column, the final corpus contains text and stances. Stop words, Punctuations and Links are removed as pre-processing steps. Either lemmatization or stemming is done. The NLTK's WordNet Lemmatizer is used for lemmatization, while theNLTK's Snowball Stemmer implementation is used for stemming, based on the Porter2 stemming algorithm. Every word is converted to lowercase in this phase to account for variances in capitalization. Hashing and TF-IDF are used under Feature extraction methods. Random Forest, Logistic Regression and Decision Tree were used. A voting classifier was also constructed to perform simple majority voting among the models' predictions</p>
--	--	--

5	Using Online Reviews for Customer Sentiment Analysis	Rae Yule Kim 12 August 2021	<p>This article explores text mining methods and proposes some alternative metrics to interpret customer sentiment. First, online review ratings have been controversial for their objectivity. Consumers who are most likely to leave a product review are either the ones who are extremely satisfied or the ones who are extremely dissatisfied. This pattern is observed from online reviews with more than 50 helpful votes, which are not likely to be potentially ingenuine “troll” or “fake” reviews.</p> <p>The volume of online reviews does not talk much about customer opinion, which matters most to ensure the future success of innovation diffusion. The advancement in text mining packages for data analytics programs such as R and Python has made the method largely accessible and provides cost-effective alternatives to popular text analysis software such as Linguistic Inquiry and</p>
---	--	--------------------------------	---

			<p>Word Count (LIWC).¹</p> <p>Text mining packages enable extracting sentiment by using lexicons such as Harvard University's General Inquirer (GI), University of Pittsburgh's Qualitative Data Analysis Program (QDAP) dictionaries with minimal effort and without charge. This paper also has information on "Using Online Reviews for Enthusiasm Analysis". Finally, the paper compared sentiment scores from various lexicon dictionaries to examine if the results are comparable and which dictionaries can be useful for marketers seeking to mine customer opinions from online reviews. this article explored the potential of lexicon-based sentiment analysis for customer opinion mining from online reviews. The paper showed that sentiment scores can be an effective alternative to measure sentiment in customer opinions because</p>
--	--	--	---

			<p>sentiment scores tend to be less influenced by the extremity bias and not skewed to extreme values. Also, the paper suggests a new angle to examine enthusiasm in customer opinion by considering the average word count in addition to sentiment scores distribution.</p>
6	<p>Textual Analysis for Online Reviews: A Polymerization Topic Sentiment Model</p>	<p>Lijuan Huang; Zixin Dou; Yongjun Hu; Raoyi Huang 30 May 2019</p>	<p>In this paper, it has been proposed a polymerization topic sentiment model (PTSM) to conduct textual analysis for online reviews. The paper applied this model to extract and filter the sentiment information from online reviews. Establishing a data dictionary, developing a sentiment prediction model, and Validating performance are the three main steps in the work. This paper develops a corpus-based method to construct a domain sentiment dictionary to solve this problem. In this paper, it is choosing the movie dataset</p>

			as the corpus. It provides a set of 50,000 highly polar movie reviews for training and testing. The PTSM model is formally presented for sentiment extraction. A sentiment prediction model is developed and a validation method is also developed. The results are gained.
7	Signal Classification and Jamming Detection in Wide-Band Radios Using Naïve Bayes Classifier	M. O. Mughal; Sunwoo Kim 27 April 2018	Radio frequency (RF) jamming is a conventional method of disrupting the communication of the targeted system. Recently, feature based algorithms for a single tone jammer detection in WB radios were proposed. This letter focuses on classifying signals and detecting a multi-tone jammer in WB spectrum sparsely populated by NB modulated signals. Compressed Sensing is one of the steps in the work. Using naive Bayes and some methods the spectral features v used for the algorithm are the maximum magnitudes of the PSD, center frequencies and

			<p>occupied bandwidths of the received NB signals. Over multiple iterations, these features are extracted and labelled with the relevant class C_k, which in this case is BPSK, QPSK or BASK. In the prediction stage, the classifier classifies the new unlabeled data. The posterior probability for every data sample to belong to a certain class is calculated</p> <p>and the class with the highest probability is assigned to the tested data sample.</p>
8	A set theory-based similarity measure for text clustering and classification	<p>Ali A. Amer & Hassan I. Abdalla</p> <p>14 September 2020</p>	<p>A similarity measure namely, similarity measure for text processing (SMTP) is proposed by Lin et al [14] In [14], the proposed measure takes the following three cases into account: a) The feature appears in both documents, b) the feature appears in only one document, and c) the feature appears in none of the documents. for knowledge discovery on text collection. While</p>

			<p>exploring the SMTP similarity measurement it is found that the case of measuring similarity between the pair of similar documents is not covered. The objective of this work is to highlight this gap and propose a minor change to make the SMTP a complete similarity measurement technique for knowledge discovery in line with the other standard similarity techniques. The case where the standard deviation for a particular feature tending to (or equal to) zero is not covered in the proposed similarity measure SMTP. The suggested change is also compared in line with the two-standard similarity measurement techniques namely Cosine and Jaccard similarity measures.</p>
9	News Recommendation Systems - Accomplishments, Challenges & Future Directions	Chong Feng; Muzammil Khan; Arif Ur Rahman; Arshad Ahmad	News publishers have shifted away from traditional newspapers and embraced digital platforms such as websites and specialized mobile apps for delivering news content.

		20 January 2020	<p>These digital platforms often employ news recommendation systems, which can automatically process lengthy articles and suggest similar ones to readers based on predefined criteria. This study aims to achieve several objectives:</p> <ol style="list-style-type: none"> 1. To identify and categorize the challenges associated with news recommendation systems. 2. To identify and categorize state-of-the-art approaches and their application domains. 3. To identify the datasets used for evaluation and their sources, along with the evaluation methods applied. 4. To explicitly highlight the challenges that have been addressed in the literature. <p>The study conducted an extensive review of research spanning from 2001 to 2019 and identified 81 relevant studies, which were broadly classified into</p>
--	--	-----------------	---

			<p>six categories for discussion. The analysis revealed that the majority (60%) of news recommendation systems adopted a hybrid approach. Furthermore, 66% of the studies provided limited information about the datasets they used, and only a subset of the numerous challenges in the news domain were addressed.</p> <p>This article represents the first comprehensive overview of the field of news recommendation, providing insights into various dimensions covered in the existing studies. In the concluding section, the paper outlines future research opportunities aimed at enhancing the recommendation of news articles in the news domain.</p>
10	Blocking reduction strategies in hierarchical text classification	A. Sun; E.-P. Lim; W.-K. Ng; J. Srivastava August 2004	<p>In this paper, it is focused on the blocking problem of methods based on this topdown level-based strategy. Documents are said to be blocked when they are wrongly rejected</p>

		<p>by the classifiers at some higher-levels and cannot be passed down to the classifiers at the lower-levels. It is clear that blocking is one of the reasons for poor recall for categories at the lower-levels. Unlike the usual category-centric measures such as precision and recall, the blocking measure is classifier-centric, as it is a value derived for a subtree classifier. The extent of blocking caused by a subtree classifier in a top-down</p> <p>HTC method can be calibrated by its blocking measure. Threshold Reduction Method (TRM), Restricted Voting Method (RVM) and Extended Multiplicative Method (EMM) are some Hierarchical Classification Methods discussed. Among them, the restricted voting method is the best in terms of FM1 measures and is able to reduce blocking significantly. However, restricted voting method</p>
--	--	--

			involves more classifiers (i.e., secondary classifiers), which require more time for training.
--	--	--	--

3. Dataset and Tool to be used (Details)

The dataset contains 20085 rows (1 heading and others data) and 4 columns. The 4 columns are 'category', 'headline', 'author' and 'short description'. The dataset will be checked for null values. The 'authors' column will be dropped. The 'category,' 'headline' and 'short description' columns will be combined into a new column called 'combined.' The words will be tokenized to get the similarity. We will use vectorization to covert textual or categorical data into numerical vectors

➤ Data Collection:

Web Scraping Libraries: Python libraries like BeautifulSoup and Scrapy are commonly used for collecting data from news websites and online sources.

APIs: Some news websites offer APIs for accessing their content programmatically, which can simplify data collection.

➤ Data Preprocessing:

Natural Language Processing (NLP) Libraries: Libraries like NLTK, spacey, and Text Blob are used for tasks like tokenization, stemming, lemmatization, and stop-word removal.

Text Cleaning Tools: Regular expressions and custom scripts can help clean the text data by removing HTML tags, special characters, and noise.

➤ Text Classification:

Machine Learning Frameworks: Libraries like Scikit-Learn and TensorFlow are used to implement and train text classification models.

Algorithms such as Support Vector Machines, Naive Bayes, and deep learning models like RNNs and Transformers can be employed.

NLP Models: Pre-trained NLP models like BERT, GPT-3, or RoBERTa can be fine-tuned for specific classification tasks.

4. Algorithms / Techniques description

Orthography is the NLP level used here. Orthography is the study of language conventions for punctuation, script and encoding. Here the sentences are tokenized to get the individual words. The tokens not containing letters are filtered out. Naïve Bayes Classification model will be used to develop a classification model to classify the news with respect to their category. Support Vector Machines (SVMs) are also used. We will also be developing a Decision Tree.

NaiveBayes Classification Model

Naive Bayes classification model is used in this news classification systems to categorize news articles into predefined topics or classes. This is applied to news classification systems for the purpose of assigning news articles to specific categories, such as politics, sports, technology, or entertainment.

The two step process includes:

1. Training: The model is trained on a labeled dataset of news articles. Each article is associated with a category label, providing the model with examples of the text content and its corresponding category.
2. Classification: Once trained, the Naive Bayes model is used to predict the category of new, unlabeled news articles. The model calculates the probability that an article belongs to each category based on the words and features within the article, and assigns the category with the highest probability as the classification for the news article.

The accuracy of this model is 63% which was low and was not up to the expectation.

Support Vector Machine(SVM) Algorithm

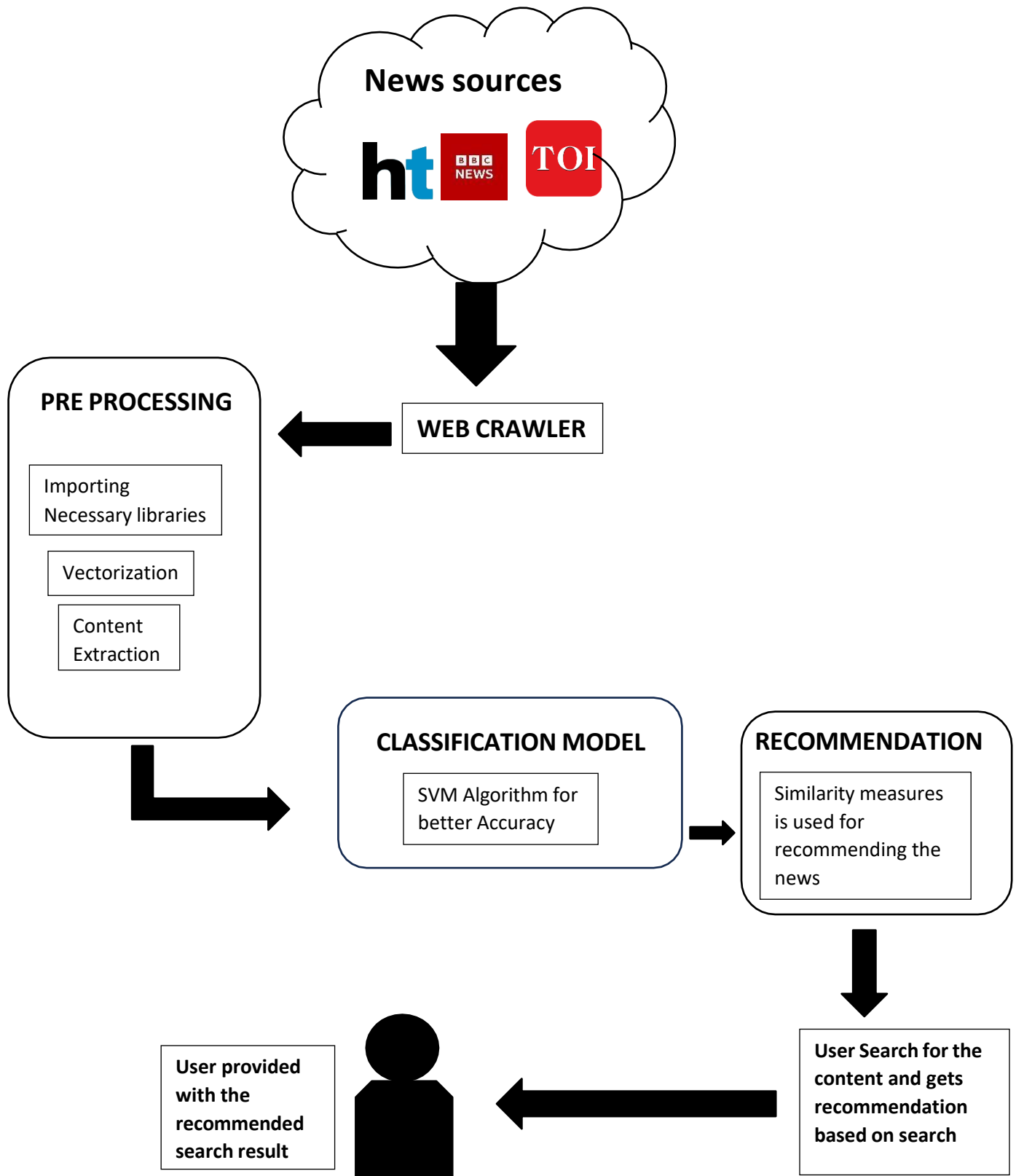
SVM algorithm was used as more accurate results were produced when compared to Naïve Bayes Classification model and we received an Accuracy of 98.2% which was more accurate with the news article dataset and the test model.

SVM serves the same purpose as Naïve Bayes algorithm but differs in the usage and the process involved are as follows

1. Feature Extraction: News articles are first preprocessed, and features are extracted from the text. Common techniques include bag-of-words or TF-IDF representations, which convert the text data into numerical vectors.
2. Training: The SVM model is trained using a labeled dataset of news articles, where each article is associated with a category label. The model learns to find the optimal hyperplane (decision boundary) that best separates the different categories based on the feature vectors.
3. Classification: Once trained, the SVM is used to predict the category of new, unlabeled news articles. The model calculates the position of a new article's feature vector relative to the learned decision boundary, and assigns the category based on which side of the boundary it falls.

The strengths of SVM in news classification systems include its ability to handle high-dimensional feature spaces, provide strong classification performance, and offer flexibility in choosing different kernel functions to capture complex relationships between features. SVMs are especially useful when there is a need for more advanced and flexible models to handle non-linear data separations in news articles.

5.ARCHITECTURE DIAGRAM



- **Bar graph representation of types of news**



The classification report consisting of the precision, recall, f1 score and support were also seen.

IMPLEMENTATION DETAILS

- Initial step involves the Data Preprocessing
- Import the required libraries
- Import the dataset to train the model so the it can be tested with the web scrapped data in real time to get the accurate results
- Check for the null characters and remove the same
- Data visualization to visualize the dataset details more clearly and check the required data drop the columns which is not required bargraph is used for the visualization and word-cloud is used to display the topics that are covered.
- Vectorization is done in order to make the classification easier by removing the stop words and required data pre processing is done.
- Classification model is done using Naïve Bayes classification model which resulted in less accurate results.
- To get more accurate results Support Vector Machine was used which provided results with higher accuracy and was considered a better solution
- Decision tree is also made for the same and is represented each of the classifications clearly
- Similarity measures are used for recommendations
- Web scrapped data can be taken in order to test the realtime data with this tested model and accurate results will be provided.

```

▶ accuracy: 0.636
┌─┐
precision  recall  f1-score  support
arts culture      0.00      0.00      0.00         6
black voices      0.83      0.42      0.56        12
business          0.00      0.00      0.00        12
comedy            1.00      0.41      0.58        32
crime             1.00      0.14      0.25         7
education         0.00      0.00      0.00         4
entertainment     0.46      1.00      0.63        58
green             0.00      0.00      0.00        13
healthy living    0.00      0.00      0.00        11
impact           0.00      0.00      0.00         8
latino voices     1.00      0.12      0.22         8
media             1.00      0.11      0.19        19
parents           0.00      0.00      0.00        12
politics          0.63      1.00      0.77       181
queer voices      0.93      0.93      0.93        30
religion          0.00      0.00      0.00         4
sports            1.00      0.29      0.44         7
style             0.00      0.00      0.00         9

```

- A Support Vector Machine model was also developed. The accuracy was 98.2% which gave more accurate results.

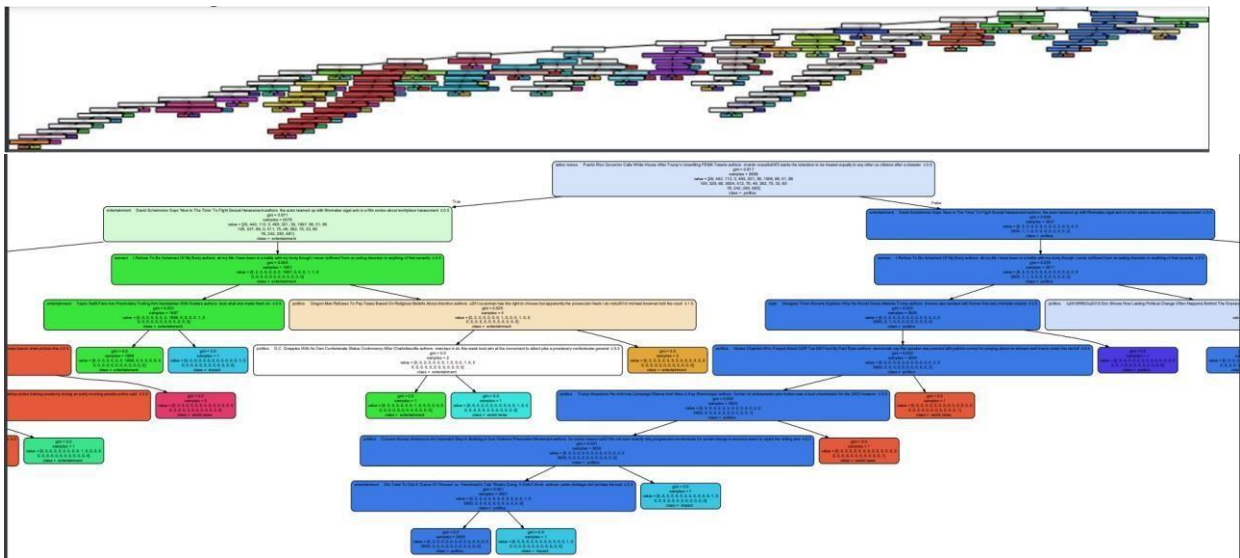
The classification report consisting of the precision, recall, f1 score and support were also seen.

```

┌─┐ accuracy: 0.982
precision  recall  f1-score  support
arts culture      1.00      0.67      0.80         6
black voices      0.92      1.00      0.96        12
business          1.00      0.83      0.91        12
comedy            1.00      1.00      1.00        32
crime             0.88      1.00      0.93         7
education         1.00      1.00      1.00         4
entertainment     1.00      1.00      1.00        58
green             1.00      1.00      1.00        13
healthy living    1.00      0.91      0.95        11
impact           1.00      1.00      1.00         8
latino voices     1.00      1.00      1.00         8
media             1.00      1.00      1.00        19
parents           1.00      0.92      0.96        12
politics          0.99      0.99      0.99       181

```

- Decision Tree



- Based on cosine similarity we can suggest similar news for a given news. Similar news can be displayed.

similar_articles

	headline	category	combined
7172	Moderate Senators Cite Progress In Talks To E...	politics	politics Moderate Senators Cite Progress ...
8598	Top GOP Races To Watch In 2018 authors	politics	politics Top GOP Races To Watch In 2018 a...
2115	McCain Undergoes Surgery To Treat Intestinal ...	politics	politics McCain Undergoes Surgery To Trea...
2128	Republicans Are Scrambling To Save An Arizona...	politics	politics Republicans Are Scrambling To Sa...
7655	A Deconstruction Of The Alt-Right Movement au...	politics	politics A Deconstruction Of The Alt-Righ...
1007	2 Senators To Watch In The Fight Over Gina Ha...	politics	politics 2 Senators To Watch In The Fight...
7071	Senator Tammy Duckworth Makes Historic Pregna...	politics	politics Senator Tammy Duckworth Makes Hi...
1282	Lawmakers In Both Parties In Talks To Save Ho...	politics	politics Lawmakers In Both Parties In Tal...
7262	With A Shutdown Looming Government Agencies A...	politics	politics With A Shutdown Looming Governme...

7. GitHub Repository Link (where your j comp project work can be seen for assessment)

https://github.com/UserYesh/WEB_MINING_DA.git

8.REFERENCES

- [1] Abdullah Tariq, Abid Mehmood, Mourad Elhadeif and Muhammad UsmanGhani Khan (2022, July). “Adversarial Training for Fake News Classification”.Vol:10.

AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/9844739>

- [2] Alaa Altheneyan and Aseel Alhadlaq (2023, March). “Big Data ML-BasedFake News Detection Using Distributed Learning”. Vol:11. AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/10078408>

- [3] Chong Feng, Muzammil Khan, Arif Ur Rahman and Arshad Ahmad (2020,Jan). “News Recommendation Systems - Accomplishments, Challenges & Future Directions”.

VOL:8. AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/8963698>

- [4] Anitha Anandhan, Liyana Shuib, Maizatul Akmar Ismail and Ghulam Mujtaba, (2018, Feb). “Social Media Recommender Systems: Review andOpen

Research Issues”. Vol: 6. AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/8303693>

[5] Rae Yule Kim. (2021, AUG).” Using Online Reviews for Customer SentimentAnalysis”.

IEEE Engineering Management Review (Volume: 49, Issue: 4, 01Fourthquarter,Dec. 2021). AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/9512387>

[6] Naveed Hussain, Hamid Turab Mirza, Ibrar Hussain, Faiza Iqbal and ImranMemon.

(2020, Mar)”Spam Review Detection Using the Linguistic and SpammerBehavioral Methods”. VOL: 8. AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/9027828>

[7] Lijuan Huang, Zixin Dou, Yongjun Hu, and Raoyi Huang .

(2019,May).“Textual Analysis

for Online Reviews: A Polymerization Topic Sentiment Model”. VOL: 7AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/8726319>

[8] M. O. Mughal and Sunwoo Kim. (2018,Apr)” Signal Classification andJamming

Detection in Wide-Band Radios Using Naïve Bayes Classifier”. IEEECommunications Letters (Volume: 22, Issue: 7, July 2018). AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/8351937>

[9] A. Sun, E.-P. Lim, W.-K. Ng, and J. Srivastava . (2004,Aug)“Blockingreduction

strategies in hierarchical text classification” IEEE Transactions on Knowledge and

Data Engineering (Volume: 16, Issue: 10, October 2004). AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/1324637> Naresh Kumar

Nagwani. (2015,Aug).A Comment on “A Similarity Measurefor Text

[10] Classification and Clustering” IEEE Transactions on Knowledge andData

Engineering (

Volume: 27, Issue: 9, 01 September 2015). AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/7177179>

[11] Juha Pärkkä, Luc Cluitmans and Miikka Ermes “Personalization Algorithm for

RealTime Activity Recognition Using PDA, Wireless Motion Bands, and Binary Decision

Tree” IEEE Transactions on Information Technology in Biomedicine(Volume: 14,

Issue: 5, September 2010) AVAILABLE:

<https://ieeexplore.ieee.org.egateway.chennai.vit.ac.in/document/5497157>

- [12] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer and R. Crompton, "Activity identification using body-mounted sensors—A review of classification techniques", *Physiol. Meas.*, vol. 30, pp. R1-R33, 2009.
- [13] V. Knnen, J. Mntyjrvi, H. Simil, J. Prkk and M. Ermes, "Automatic feature selection for context recognition in mobile phones", *Pervasive Mobile Comput.*, vol. 6, pp. 181-197, 2010.
- [14] Y.-S. Lin, J.-Y. Jiang and S.-J. Lee, "A similarity measure for text classification and clustering", *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1575-1590, Jul.2014.
- [15] P. H. A. Faustini and T. F. Cov oes, "Fake news detection in multiple platforms and languages", *Expert Syst. Appl.*, vol. 158, Nov. 20