International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 07 – Issue 03 || March 2022 || PP. 44-52

Machine Learning Based Analysis of Media News for its Content Accuracy

Nida Khan[1] and Megha Gupta[2*]
[1]*Department of Computer Application, J.C. Bose University of Science and Technology, YMCA, Delhi, India*
[2]*Department of Computer Science, Mata Sundri College for Women, University of Delhi, India*

Abstract: Media plays a significant role in today's world. It directly or indirectly affects us and also influences us. Now when we look back, we can easily find the broad evolution of change that took place in the industry of communication i.e., media and mass communication. We keep ourselves in the up in the race only if we end up having a great network and best channel that gives update about a fact with pace and precision. In this hustle of providing fact, a continuous trade between precision and pace being fought in cold manner. Like today there are so many channels available on social networking sites to provide information. But to find out which one is worthy the most. Media is really influential in recent times. As news in market are vividly in market of both type like fake and real. Fake news most of the time have poor impact on public. Like people are moved by news majorly. Not all but still little channels are trying to publish the real ground report. Real media really assures to match standard of today's mass media and communication. To understand this fact better there are several machine learning techniques available for analysis of such channels. These algorithms come out with a wider understanding over the datasets. Results depends upon dataset recorded and target class, an algorithm is applied on dataset. Algorithm guides better in field of research and improves the understanding of the dataset. This paper tries to find the role of media in the accuracy of news. This paper finds the accuracy and confusion matrix and further derived classification report of the models to predict the fake news broadcasted by media.
Keywords: Machine Learning, Fake News, Mass media, Accuracy

Introduction

Communication refers to transfer of data from one person to another through a medium. In every walk of life today media plays an inescapable role. Media means various modes of communication. Media is often known as mass media, when communication is done for large group. Medium associated with mass communication are mainly television, radio, newspaper and most commonly used is internet. Internet is basically connecting all the major domains of mass media.

Mass media is majorly influencing all major segments of society. Mass media is really convincing us in our thinking ability. Our inclination towards media is vast. So, mass media plays a vital role in life of an individual. It keeps us connected with entire nation. Everyone wants to be updated. Updated in the mean sense most recently and appropriately updated. So here comes the role of various newspaper and television news.

For democratic countries media is very necessary as our lives mainly resolve around the media. Every bit change affect are life either slightly or majorly. But influence of media do-does matter. It is certainly very crucial to know that anyway data getting shared is really factual or bias. In recent time, inclination towards media have increased. Due to outbreak of global pandemic, our reliability over media for information is increased. Gadgets plays a vital role in our connection with social networking site. As advancing technology is growing so is demand for appropriate information. Every information providing software works constantly to provide most updated knowledge. These software gets frequently updated. Also, nowadays people have shifted their ways of gathering news details. Now every news channel has well organized platform on all usual websites. So, trends of proneness over different sites data can be gathered.

Media have been predominant in history of recent times. Like for better understanding of ground reality of any instance it's common to rely on easily available stuff. Real news are important for us. Fake news are really affecting lives of people. Reason behind the same is dominant role of media. In every walk of life, people are getting blown with this hustle of media. Now, mass media is used getting spread through the various medium. With change in time individuals are finding new ways of playing with media. In recent times, almost all the famous news channel are having their account on all the platforms. News on all media is easily avail by individual. The main trending trade on social media in account with mass media and news channel is trade of providing current news. This is duty of concerned news authority to get factual content instead of misguiding content. Fake news often end up building unwanted mess. In some senses, it is duty of us as user to try to segregate between real and fake news. Various research have been done on datasets to understand and focus on how attribute of news channel are leading to better understanding of data.

There are several machine learning approaches to analyze a particular dataset. We are basically going to implement given below approaches to understand media dataset:

- Decision tree- A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label.
- K-Nearest Neighbor (KNN) Algorithm- It is a supervised learning technique which assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- Logistic regression- Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.
- Naïve Bayes- Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable.
- Random Forest - It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.
- Support vector machine- Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points.

In this paper we will be taking up a media news dataset and apply machine learning algorithm to understand a number of facts are associated with media. The dataset holds target class as factual reporting rating and testing set as bias rating. Dataset also have an attribute stating names of various site available as media. We are using python as support tool for implementing machine learning algorithms. We will be applying machine learning algorithm on our dataset. Machine learning approaches are predefined in python by importing required libraries easily algorithm can perform on any dataset. Here, machine learning procedure executed namely are:- decision tree, KNN, random forest, SVM, naïve bayes and logistic regression. Next, deriving various models and precision associated with same. Confusion matrix after every algorithm is recorded. Later graphs of prediction are implemented.

Related Work

The work (Hamborg et al. 2019) gives literature review to show area of media bias through automated identification. In (Aggarwal et al. 2020), study tests media subjectivity versus objectivity by examining news events reported from their Twitter accounts. Research then shows how such subjective articles program news consumers and regulate their opinions. Finally, a support system that detects alarming biases in the media is provided through a unique mechanism for calculating short-term impact bias scores.

The book "Mass Communication in India" (Kumar 2020), shares that the fourth industrial revolution is hereby shifting. Also tells about the emergence and advancement of each of the upcoming industrial revolutions. The author in work (Gladstone & Neufeld (2012), majorly provides information about "a treatise on our connection with the news media". It also explains American history of media.

Skewed: A Critical Thinker's Guide to Media Bias by Larry Atkins 2016 is a wonderful piece highlighting most important fact that in world of competing and controversial media, how a common man gets aware with true fact. Author also shares about role of media in separation of factual facts and agenda driven campaigns.

Schiffer's ideal volume on the genuine issues with news media exposes normal claims of political inclination, specifically liberal predisposition. The creator clarifies in his presentation that the media here and there incur genuine injury for educated citizenship through a reiteration regarding schedules, predispositions, and inadequacies that leave news customers unprepared to explore contemporary legislative issues (Schiffer 2017).

The (Park 2016), review research gives the idea and nature of journalistic prejudice and gives proposals for those in the media to establish a more straightforward climate for writers and the overall population. This paper endeavors to cure this issue, which may before long prompt considerably more prominent disappointment, maybe even disdain, if the inescapable view of journalistic spin isn't amended.

Khanam et al. (2021) makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised machine learning algorithms that can classify fake news as real or fake by text analysis.

The Guess et al. (2020) shares about beginning around 2016, there has been a blast of interest in falsehood and its part in decisions. Research by media sources, government organizations, and scholastics the same has shown that huge number of Americans have been presented to questionable political news on the web. Notwithstanding, generally little examination has zeroed in on reporting the impacts of burning-through.

Fake News and the Third-Person Effect: They are More Influenced than me and you (Ștefăniță et al. 2018) displays how fake news and third person effects individual's opinion over the news shared on media. Late exploration endeavors have been put into sabotaging the impacts of advanced disinformation, both on an individual and on a cultural level. Notwithstanding, in light of the intricacy of the peculiarities, the genuine impacts of advanced disinformation are as yet getting looked at and, in this manner, studies distributed so far center around the apparent impacts of phony news.

The work (McChesney 1993), shows in detail the emergence and consolidation of U.S. commercial broadcasting economically, politically, and ideologically. This process was met by organized opposition and a general level of public antipathy that has been almost entirely overlooked by previous scholarship.

Leighton Andrews in 2019 puts about Facebook, the Media and Democracy analyzes Facebook Inc. furthermore the effect that it has had and keeps on having on media and majority rule government all over the planet.

Another work "Third person effects of fake news: Fake news regulation and media literacy interventions" 2018 talks about how fake news and media influences the society and individual in day-to-day life. Analysis of certain data is being done to understand how third person effects individuals to practice on news report.

The theoretical structure created in the review (Hallin & Mancini 2004) ended up being a significant contribution to the field of the near media frameworks research since it gives a precise and pertinent way to deal with investigate contrasts and similitudes of the connections among media and legislative issues.

Ravi in 2017 gives work on Current Media, Elections and Democracy investigates how the cutting-edge media capacities in a popular government, particularly during decisions, when it plays out the essential job of teaching individuals and embellishment general assessment. At such critical points in time, a field for public discussion and some of the time even a check against the maltreatment of force.

Gil et al. (2018) said that as the second quarter of the 21st century approaches, the study of social media and its influence on democracy has rapidly penetrated into different areas of the social sciences, especially communications. politics. Building on the evidence accumulated in this body of literature, this paper briefly summarizes several established areas of research.

A book by James Curran 2011, Media and Democracy gives an opinion about relation between media and democratic world. More is discussed about major problems faced by media. It also talks about role of media in political situation.

Entertaining Democracy book by James P. Curran gives abstract view on Mass Communication and Society is a title that is well known to students worldwide for its ability to deliver insightful and accessible essays leading international scholars on the most relevant issues in today's media (Curran 2010).

Methods

In field of artificial intelligence, one of most blooming area is machine learning. It implements an algorithm or method to extract patterns from raw data. The goal of machine learning is to allow systems to learn from their experiences. It works without any explicitly programmed methods, also human involvement is not much needed. Machine learning basically involves 3 steps to work on a given dataset.

The steps necessary to generate models of any algorithms as shown in figure1 are:

1. Collection of dataset
2. Operating on dataset
3. Evaluation of results

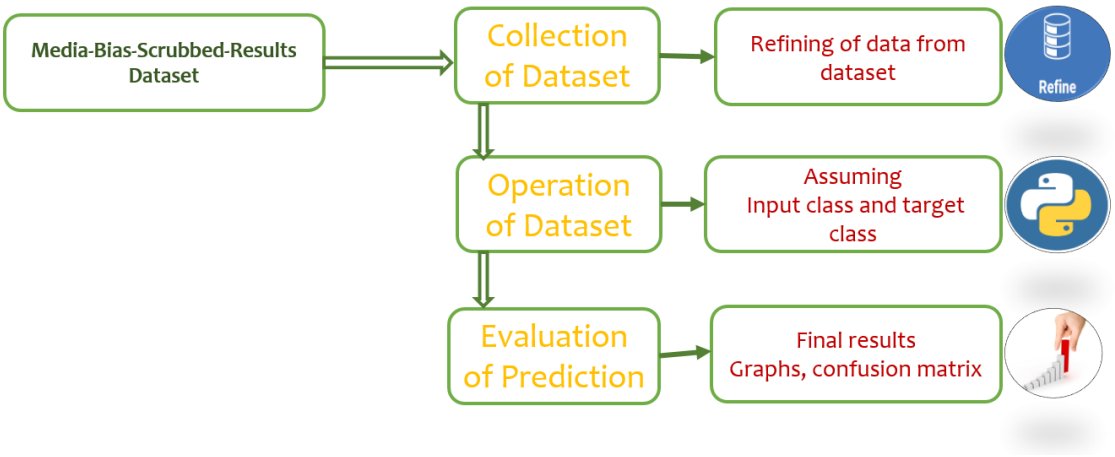

Figure 1. Steps to apply on machine learning algorithm

For preparing models of any machine learning algorithm the very first step is collection of data. Dataset is firstly refined and then used with various classifier. Data need to be tuned basically before usage.

Operation on dataset is most important dataset as in this step we classify dataset attributes as input class and predicted or target class. Input class and predicted class are split into any ratio as training and testing set. The algorithms are applied on input variable and then predictions are made on target class. Algorithms of machine learning are implemented over dataset.

After codes are executed, prediction is presented and also accuracy of model used by algorithm is evaluated. Confusion matrix is also generated on every model of classifier. Graph of predicted output is also displayed. Testing data points, training data points and predicted target class is also plotted and visualized using graph.

Evaluation

Using python as language for research given below information was extracted about the data by applying machine learning models. There are many pre-defined libraries in python. In order to evaluate data here, we are taking our dataset and firstly considering it's one of the attributes as input class variable(X) and other as target class(Y). We are then splitting our data set into training set and testing set. After this we are applying some machine learning algorithms. Above all we are creating classifier models. Given below are list of algorithms we imposed on our dataset. Figure 2 shows dataset: media-bias-scrubbed-results A fully scrubbed CSV of all of media bias fact check's primary categories (note on bias negative (-) connotes liberal bias, positive (+) connotes conservative bias) (Media CSV file).

	site_name	url	bias_rating	factual_re
2	(The) Interpreter Magazine	http://www.interpretermag.com/	-8	HIGH
3	1600 Daily	https://www.whitehouse.gov/1600daily	26	MIXED
4	2ndVote	https://2ndvote.com/	29	MIXED
5	680 News	http://www.680news.com/	-3	HIGH
6	71 Republic	https://71republic.com/	15	HIGH
7	9 News (Australia)	http://www.9news.com.au/	6	HIGH
8	ABC News Australia	http://www.abc.net.au/news/	-4	HIGH
9	ABC News	http://abcnews.go.com/	-13	HIGH
10	ABC11 Eyewitness News	http://abc11.com/	-11	HIGH
11	Above the Law	http://abovethelaw.com/	-9	HIGH
12	ABS-CBN News	http://news.abs-cbn.com/	-5	HIGH
13	Acculturated	https://acculturated.com	24	HIGH
14	Accuracy in Academia (AIA)	https://www.academia.org	28	MIXED
15	Accuracy in Media (AIM)	http://www.aim.org/	30	MIXED

(a)

	site_name	...	factual_reporting_rating
0	(The) Interpreter Magazine	...	HIGH
1	1600 Daily	...	MIXED
2	2ndVote	...	MIXED
3	680 News	...	HIGH
4	71 Republic	...	HIGH
...	...	...	...
1599	Young Conservatives	...	MIXED
1600	Your Black World	...	MIXED
1601	Youth Radio	...	HIGH
1602	Z Magazine	...	HIGH
1603	ZDF (Zweites Deutsches Fernsehen)	...	HIGH

[1604 rows x 4 columns]

(b)

Figure 2. (a)Dataset: media-bias-scrubbed-results (b) Data frame using python

In this dataset we are considering bias rating as input class and factual reporting rating as target class. Model of given below machine learning techniques is used:

- Decision tree
- K-Nearest Neighbor (KNN) Algorithm
- Logistic regression
- Naïve Bayes
- Random Forest
- Support vector machine

Given below is list of classifiers used for creating models of classifier:

- Decision tree – Decision Tree Classifier () is provided by
- K-Nearest Neighbor (KNN) Algorithm
- Logistic regression
- Naïve Bayes
- Random Forest
- Support vector machine

Figure 3. Heat map of the dataset.

Dataset got split into two training set and testing set as 70% and 30% respectively. Same split was followed by all the models. Figure 3 shows the heat map of dataset. The accuracy for same was recorded. Confusion matrix was also generated. Prediction was also displayed after every model calculation.

Given below are results after performing all the above-mentioned machine learning algorithm on the data set.

Score is measuring the accuracy of the model against the training data. Table 1 exhibits the accuracy score of all the models on training set and testing set.

Result and Discussion

Machine learning model	Accuracy of training set(%)	Accuracy of testing set(%)
Decision tree	84	83
K-Nearest Neighbor(KNN) Algorithm	84	84
Logistic regression	81	79
Naïve Bayes	80	80
Random Forest	84	84
Support vector machine	70	70

Table 1. Accuracy score on training set and testing set

Machine learning model	Classification Accuracy (%)
Decision tree	82.98
K-Nearest Neighbor(KNN) Algorithm	83.81
Logistic regression	78.81
Naïve Bayes	80.20
Random Forest	84.02
Support vector machine	70.33

Table 2. Accuracy score on predicted set

Classification Accuracy is the ratio of number of correct predictions to the total number of input samples. Table 2 exhibits the Accuracy of all the models on predicted set.

Classification report was also generated after preparation of all the models. Using classification report F1 score and precision of models were gathered. Now, here is table 3 lay out F1 score and precision with respect to target class variable of all the models.

Machine learning model	F1 Score(%)			Precision(%)		
	High	Mixed	Very high	High	Mixed	Very high
Decision tree	89	73	0	83	83	0
K-Nearest Neighbor(KNN) Algorithm	89	76	0	85	81	0
Logistic regression	86	61	0	78	85	0
Naïve Bayes	87	65	0	80	80	0
Random Forest	89	76	0	84	84	0
Support vector machine	83	0	0	70	0	0

Table 3. Classification report of model

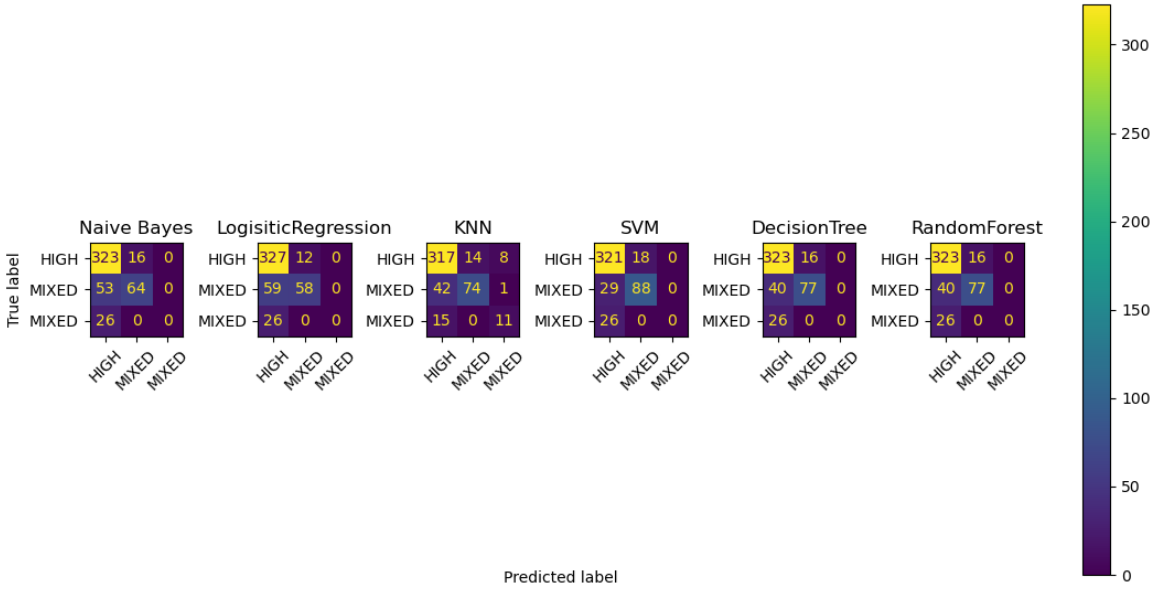

Figure 4. Confusion matrices of model

Confusion matrix provides the description of classifiers whose true value are known. Confusion matrix of all classifiers is put on show by figure 4. Pictorial representation of predicted output is best way to express dataset. Dataset and predicted output are scatter placed over the graph. Figure 5 displays scatter plot of model with highest accuracy. Random forest gives the maximum accuracy. Figure 6 displays scatter plot of model with lowest accuracy. Support vector machine (SVM) model gave the worst accuracy on the dataset.

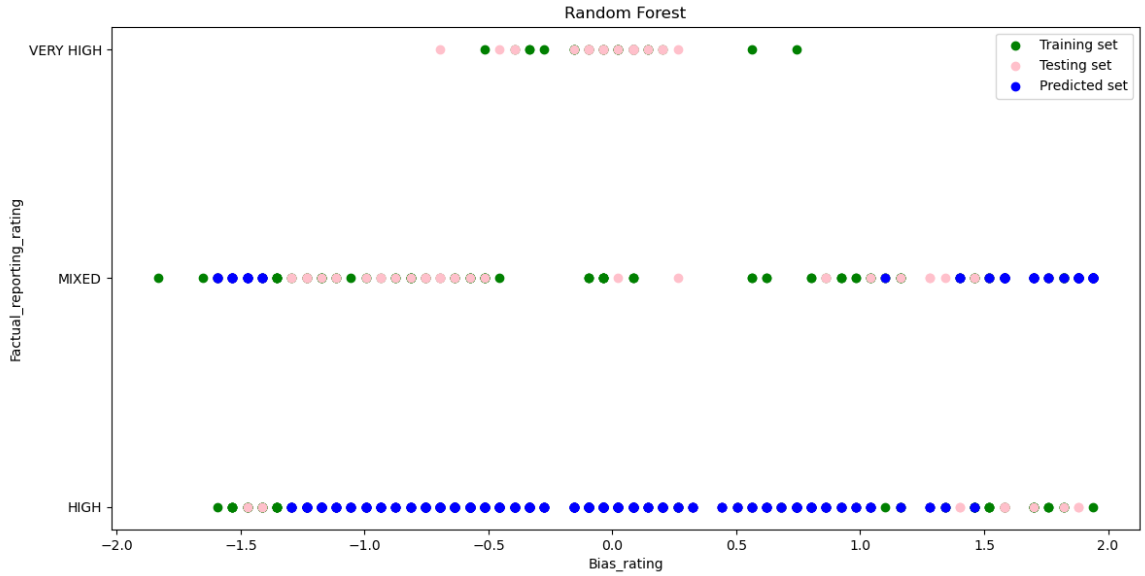

Figure 5. Scatter plot of Random forest model

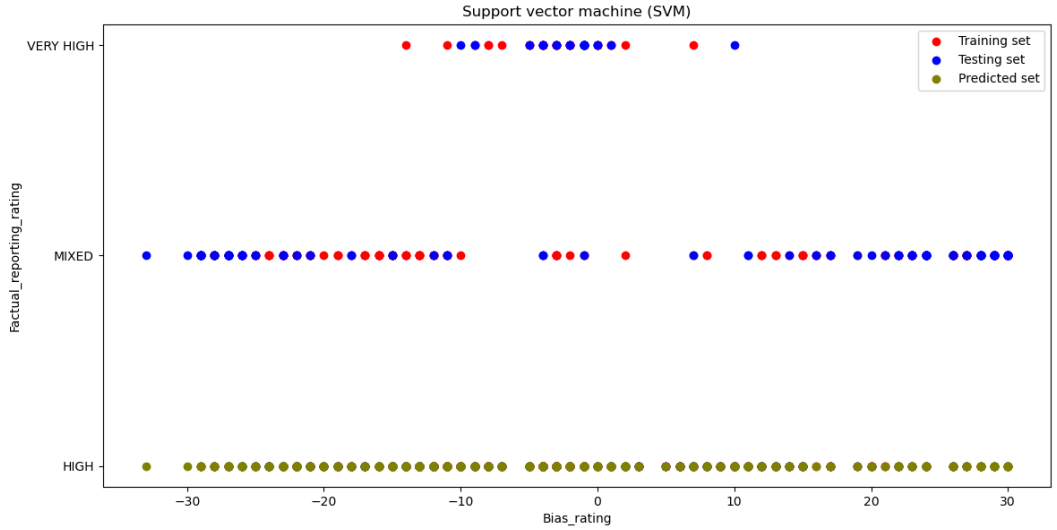

Figure 6. Scatter plot of SVM model

Using above result we can conclude that machine learning techniques are really beneficial for us to draw conclusions. As above various machine learning techniques were applied over the dataset and results were captured.

On summarizing the entire above results we wind up with following information:

- For training set i.e., 70% of dataset worked best for on decision tree, K-Nearest Neighbor (KNN) and random forest with accuracy of 84%.
- For testing set which is 30% of original dataset accuracy of 84% was achieved on K-Nearest Neighbor (KNN) and random forest.
- Accuracy on predicted dataset made best 84.02% on random forest model. Lowest accuracy was for simple vector support model of 70.33%.
- Precision came out to be 85% on K-Nearest Neighbor (KNN), 85% on logistic regression and 0% on all models for target class high, mixed and very high on dataset.
- F1 score turned out to be maximum of about 89% on decision tree, K-Nearest Neighbor (KNN) and random forest on target class named high. It was 0 on all models of target attribute, very high. On mixed class maximum F1 score of 76% was on K-Nearest Neighbor (KNN) and random forest.

The above information was gathered by applying various machine learning models. Machine learning algorithm gives best precision on suitable dataset. Graphical representation of dataset is really beneficial. More work can be done this dataset. More visualization can be done for better understanding of datasets. Various results are derived like which algorithm was best for the dataset in terms of precision and accuracy. F1 score was also derived using classification report.

Conclusions

Media plays a vital role in day-to-day affair of life. Media have changed its form in recent times. In this paper, different models were prepared in order to understand the dataset of media. Role of media is also explained. Results were derived in various forms. Total of 7 models are prepared for the dataset. Confusion matrix of the dataset was also prepared. Results are represented in tabular form and also, classification report was also presented. Graph of the model with highest accuracy is placed in the paper. Media dataset holds 3 target class variables very high, high and mixed. Confusion matrix gave more visual idea of prediction. F1 score of all the model is also prepared in the paper. Future work needs to be done in order to understand the media in much better way. More data can be analyzed to understand role of media. As forms of media is changing so is way of understanding media. Target class can be improved in order to understand the set better. More work can be done to detect impact of fake news. Role of appropriate news can also be analyzed. More input variables can get associated for better understanding of media.

Reference

[1]. Aggarwal S., Sinha T., Kukreti Y. & Shikhar S.(2020) Media bias detection and bias short term impact assessment, Array, vol 6, 100025.
[2]. Andrews L. (2019) Facebook, the Media and Democracy: Big Tech, Small State?
[3]. Atkins L. (2016) Skewed: A Critical Thinker's Guide to Media Bias Hardcover
[4]. Curran J. (2011) Media and Democracy, Taylor Francis
[5]. Curran, James P. (2010) Entertaining Democracy. In: James P. Curran, ed. Mass Media and Society, Fifth Edition. London: Bloomsbury, pp. 38-62.
[6]. Gladstone B., Neufeld J. (2012) The Influencing Machine – Brooke Gladstone on the Media, Paperback Edition
[7]. Gil de Zúñiga, Homero & Huber, Brigitte & Strauss, Nadine. (2018). Social media and democracy. El Profesional de la Información. 27. 1172. 10.3145/epi.2018.nov.01.
[8]. Guess, AM; Lockett, D; Lyons, B., Montgomery M. J., Nyhan B., Reifler J.(2020) "Fake news" may have limited effects beyond increasing beliefs in false claims.
[9]. Hallin C. D., Mancini P. (2004) Comparing Media Systems: Three Models of Media and Politics
[10]. Hamborg, F., Donnay, K. & Gipp, B. (2019) Automated identification of media bias in news articles: an interdisciplinary literature review. Int J Digit Libr 20, 391–415.
[11]. Jang M. S., Kim K. J., (2018) Third person effects of fake news: Fake news regulation and media literacy interventions.
[12]. Khanam Z., Alwasel N. B., Sirafi H. and Rashid M. (2021) Fake News Detection Using Machine Learning Approaches, IOP Conf. Ser.: Mater. Sci. Eng. 1099 012040.
[13]. Kumar J. K. (2020) Mass Communication in India, PaperBack Fifth Edition
[14]. McChesney. W. R. (1993) Telecommunications, Mass Media, and Democracy.
[15]. Media-bias-scrubbed-results.csv, https://gist.github.com/nsfyn55/605783ac8de36f361fb10ef187272113
[16]. Park D. (2016), Media Bias: How the bias affects public perceptions of the media and what can be done to further prevent erosion of media- public relationship, A Senior Project presented to The Faculty of the Journalism Department California Polytechnic State University, San Luis Obispo.
[17]. Ravi K. B. (2017) Modern Media, Elections And Democracy
[18]. Schiffer J. A. (2017) Evaluating Media Bias
[19]. Ștefăniță O., Corbu N., Buturoiu R. (2018) Fake News and the Third-Person Effect: They are More Influenced than Me and You, Journal of Media Research, vol. 11, 3(32).