FAKE NEWS DETECTION USING GATs

Aditya Chayapathy
achayapa@asu.edu

Arun Karthick Manickam
amanick4@asu.edu

Jagdeesh Basavaraju
jbasavar@asu.edu

Anuhya Sai Nudurupati
anuduru1@asu.edu

Abhijith Shreesh
ashreesh@asu.edu

ABSTRACT

The average news consumption via non-traditional sources such as social media, blogs and instant messaging groups have certainly gained a lot of interest in the recent time. One of the main sources of news content that people rely nowadays upon is via social media. Upon the release of a new data source, there are a set of issues associated with it. The rise of social media as a news source has the issue of generation and widespread of falsified information. This poses a threat not only specific to a platform in which the news content is published but also to the Internet as a source of medium for communicating news as a whole. In this project, we are trying to implement a proactive methodology to predict fake news that prevalent in social media. The essential idea behind the system is to exploit the social relationship in social media to build a semi-supervised detection system.

Keywords: *Social Media; Graph Attention Networks; BERT; User-Modeling; Social Media Mining; Data Cleansing; Text Modelling; Word Vector;*

1. INTRODUCTION

The usage of the web as a medium for perceiving information is increasing daily. The amount of information loaded in the social media at any point is enormous, posing a challenge to the validation of the truthfulness of the information. The main reason that drives this framework is that on an average 62% of US adults rely on social media as their main source of news. The quality of news that is being generated in social media has substantially reduced over the years.

The generation of fake news is intentional by the unknown sources which trivial and there are existing methodologies to individually validate the users' trustworthiness, the truthfulness of the news and user engagement in social media. But, analysing these features individually doesn’t consider the holistic factors of measuring the news credibility. Hence, combining the auxiliary information together with the news content to measure the news credibility is a possible route to focus. There have been techniques to validate the writing style of the users to classify the news content but these methods also have their outliers and error rates.

2. PROBLEM STATEMENT

The problem statement of this research revolves around implementing a system which would proactively measure the truthfulness of the news content that is published in the social media. The necessity of the system is to truncate the rapid spread of the falsified information that pose a threat to social media as a platform. In this project, we are striving to leverage auxiliary information available in social media to classify the news content. The main factors that are involved in the social media news ecosystem are the news articles and the users (publishers, readers etc). In this ecosystem, once the news content is published, the news is not only validated against the authenticated sources, but also the other users engagement, the publisher bias towards the published news topic and the historical credibility values of the user is taken into consideration to a more justifiable classification of the news content. On a high level, this system have computes the News-News relationship from the social media using the User-News relationship and derives the News Content Vector using the BERT Library, which is in-turn fed into Graph Attention Networks (GATs) scheme for prediction. The machine learning model here is based on semi-supervised technique which continuously learns the volatile social relationships and prone to change user-credibility score before prediction.

3. RELATED WORKS

a. Unsupervised Fake News Detection on Social Media: A Generative Approach, Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, Huan Liu; 2019 Association for the Advancement of Artificial Intelligence (www.aaai.org).

   Summary:

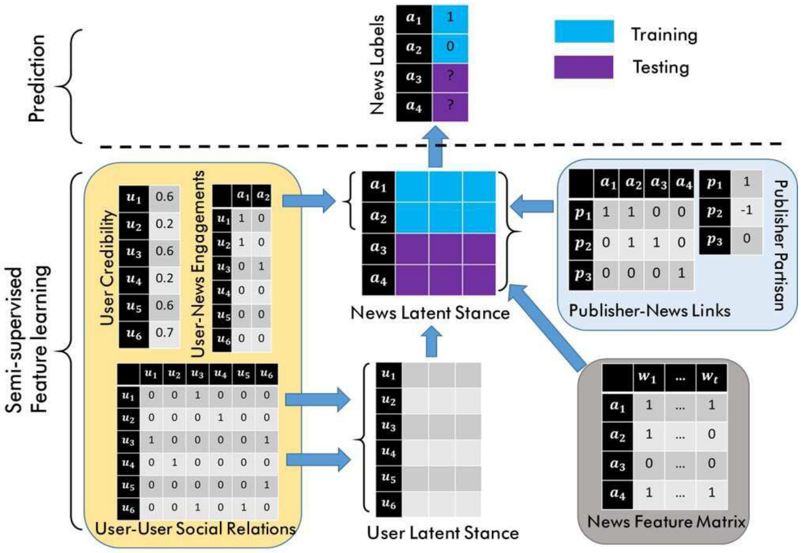
   Most of the current research in the field of fake news detection use supervised means to classify the news as either fake or real. This involves training a classification model based on a predefined labelled dataset. Though these methods show significant results, they suffer a major drawback i.e. they require a reliably annotated dataset to train a classification model. This process is time-consuming and labour intensive as it requires careful checking of news contents as well as other additional evidence such as authoritative reports. In order to mitigate this issue, the authors of the paper propose an unsupervised learning framework, namely UDF, to identify the fake news. The key idea involves extracting users' opinions on the news by exploiting the auxiliary information of the users' engagements with the news tweets on social media and aggregating their opinions in a well-designed unsupervised way to generate the final estimation results. UDF first extracts the users' opinions on the news by analyzing their engagements on social media and builds a Bayesian probability graphical model capturing the complete generative process of the truths of news and the users' opinions. Next, an efficient collapsed Gibbs sampling approach is proposed to detect fake news and estimate the users' credibility simultaneously.

b. Exploiting Tri-Relationship for Fake News Detection; Kai Shu, Suhang Wang and Huan Liu; 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org).

   Summary:

   The idea behind this paper is to leverage the three auxiliary information available in the social media regarding the news, publisher and engagers to effectively classify the news content. The tri-relationship is established between news publisher, news and social media users. The framework discussed in this paper is based on semi-supervised machine learning classification technique where the three matrices namely User-User Social Relationship, User Credibility and New User Engagement are the basic requirements. Initially, the matrix values are are inputted based on the current social media content, as and when the user relationship and engagement changes, the values are updated parallely for better classification. In this way, the model is continuous learning the social relationship and user credibility values. Once the model is prepared, the Publisher News Links and News Feature matrix are the test parameters to predict the News truthfulness. This method also considers the user bias which is

Publisher Partisan metric towards the topic/entity the news content is based on, which significantly improves the performance of the framework. When it comes to performance and accuracy compared to other traditional methods, TriFN achieves average relative improvement of 9.23%, 8.48% on BuzzFeed and 6.94%, 8.24% on PolitiFact, comparing with LIWC+Castillo. Also, issues like cold start problem are generically handled by introducing delays in the computation and modelling.



News Labels	
a_1	1
a_2	0
a_3	?
a_4	?

User Credibility	
u_1	0.6
u_2	0.2
u_3	0.6
u_4	0.2
u_5	0.6
u_6	0.7

	a_1	a_2
u_1	1	0
u_2	1	0
u_3	0	1
u_4	0	0
u_5	0	0
u_6	0	0

	u_1	u_2	u_3	u_4	u_5	u_6
u_1	0	0	1	0	0	0
u_2	0	0	0	1	0	0
u_3	1	0	0	0	0	1
u_4	0	1	0	0	0	0
u_5	0	0	0	0	0	1
u_6	0	0	1	0	1	0

	a_1	a_2	a_3	a_4
p_1	1	1	0	0
p_2	0	1	1	0
p_3	0	0	0	1

Publisher Partisan	
p_1	1
p_2	-1
p_3	0

	w_1	...	w_t
a_1	1	...	1
a_2	1	...	0
a_3	0	...	0
a_4	1	...	1

Fig 1: Tri-Fn Model

4. SYSTEM ARCHITECTURE & ALGORITHMS

As mentioned in the problem statement, the project deals with identifying fake news from the given dataset of BuzzFeed and PolitiFact. The implementation involves tasks such as data preprocessing, feature extraction, training models etc. The architecture diagram of the implementation is provided in *fig 2*.

The detailed explanation of each of the steps in the diagram is provided below.

a. Data Preprocessing:

The data was obtained from the DMML lab at ASU. The dataset comprised of information (in the form of JSON files) of news articles from BuzzFeed and PolitiFact. Along with this, the dataset provided information about users and user-news interaction in the form of CSV files. Furthermore, the dataset included real and fake news content. Based on our intuition and research, we found that the “body” of the news articles, i.e, main content, best represents the news articles core information and variance. After extracting the body of all the news articles we constructed a data frame of the news-ids and the body of the text. A label was also added to the data frame to indicate whether the news article is fake or real.

Hence, each row/document was a news article with its content and a label. This data frame was constructed independently for BuzzFeed and PolitiFact.

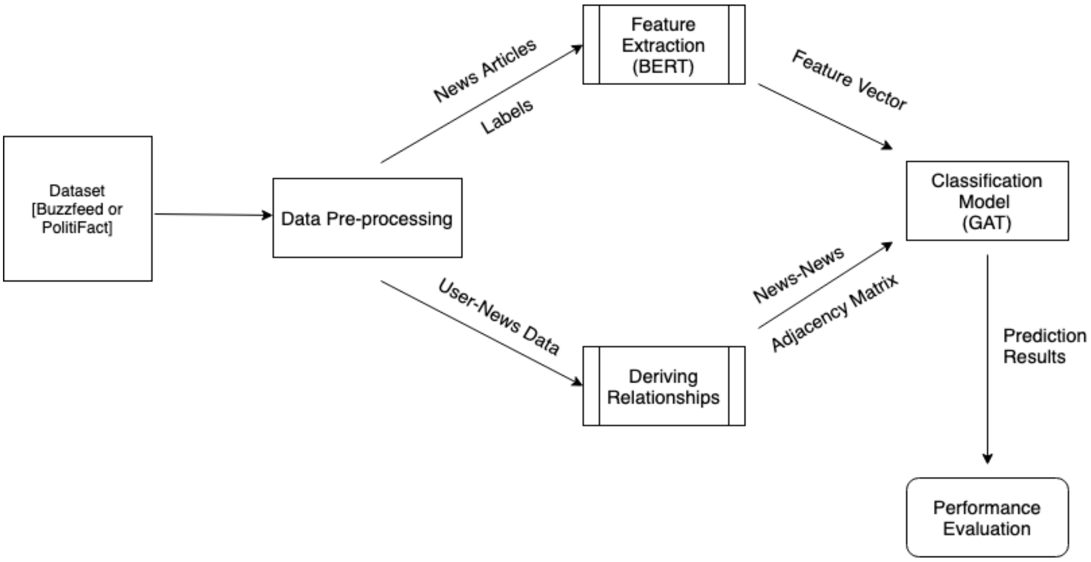


Fig 2: Architecture Diagram

b. Derive Relationships:

Once the important aspects of the data were identified for feature extraction, the next step involved establishing relationships among the news articles that can be leveraged by classification model. The relationship between the news articles was established using a graph data structure. Here, the news articles act as the nodes of the graph and the edges between the nodes represent the relationships among the new articles. The edges between the nodes of the graphs were established using the following steps:

1. Identify all unique news articles (nodes).
2. For each news article, identify the users associated with it.
3. For each of the users obtained in step 2, identify the news articles associated with them.
4. For each of the news articles obtained in step 3, establish an edge to the initial news article (step 2).
5. Represent the resulting graph in the form of an adjacency matrix. The resultant adjacency matrix is used for classification.

c. Feature Extraction:

For the fake news detection, the actual news data (body of the news article) is being considered as features. But the data is in the form of text. It is known that for the machine learning analysis, text data does not work well. So the text data has to be converted into a numerical representation. This process is called vectorization. Every record (i.e news article in this case) should be converted into a vector. There are several techniques/algorithms which can convert text to vector. Below is the list of such techniques:

1. Bag of Words (BoW) model

   This is one of the traditional models used to convert text to numbers. In a given set of records, certain words are identified as significant words. The models generally have these prebuilt set of words. Consider the count of such words equal to ‘n’. So the vector generated for the word will be of length n (number of components). If the ith word is present in the document (or text), then the ith component in the corresponding vector will 1, 0 otherwise. This way the vectors are generated for all the documents (or articles). This is the simplest vectorization technique.
   The main challenge in this technique is that all the words (among the ones you chose) are weighted uniformly which is not true in all scenarios as the importance of the word differs with respect to context.

2. TF-IDF

   TF-IDF stands for Term Frequency-Inverse Document Frequency. In this model, the words are assigned a weight based on the frequency of appearance. The model has 2 parameters as mentioned in the name. The term frequency component adjusts the weight proportionally with the number of times the word appears in the document with respect to the total number of words in that document. Inverse document frequency component identifies unique words in the set of documents and increases weight accordingly. If a particular word is appearing in most of the documents, then its weight is reduced as it will not help anyway in distinguishing the documents. Though this model weights the words based on the frequency and unique factors, it is not able to capture the meaning of the word.

3. Word2Vec

   In the models explained till now, the context of the word is not taken into consideration. The same word appearing at two different locations in the same text convey different meanings. Word2Vec is a vectorization model which takes this context into consideration. This model takes the surrounding words into context to generate the target word. This is a neural network driven model and is an improvement over the previous models. Word2Vec models take local context into consideration i.e the surrounding of the word. But it fails to recognize the global context.

4. BERT

   BERT stands for Bidirectional Encoder Representations from Transformers [4]. It is one of the state-of-the-art vectorization techniques which has achieved state of the art results on the standard datasets. This is also driven by a neural network. As the name suggests, the technique masks a particular word and tries to predict the word by running a neural network from both sides (Forward and backward). BERT is the first successful model to implement bidirectional vectorization. As a result of these fantastic results achieved by this model, in this project, BERT was used to obtain feature vectors from the text data (news article).

BERT has a pre-trained model (based on their pre-trained corpus) exposed for the users to provide the text or document and get back the vector which can be used for further analysis. BERT also gives the flexibility of building the model from scratch by providing the corpus. But in this project, the pre-trained model has been used as the content of the news articles is not too specific for the model to be trained from scratch. The news articles are fed into this pre-trained model and the resultant vectors are used as feature vectors for further analysis.

d. Training Classification Model:

As explained in the previous sections, the adjacency matrix, feature vectors and the labels form the input for the classification model. The fake news detection task has network-based input in the form of adjacency matrix representing the relationship between news articles. As explained before, this relationship is derived from the information provided in the form of users and their association with the news articles (post, share, retweet etc). As one can observe, the data has the pattern of a graph with nodes or vertices being the news articles and edges as the relationship between them. The classification task needed a model which can utilize this network or graph-based architecture of the dataset and the feature vectors generated for the news articles using BERT. Traditional machine learning models utilize just the feature vectors for prediction and will not leverage the graph-based architecture of the data. The solution to this problem was obtained through Graph Attention Networks (GATs) [2].

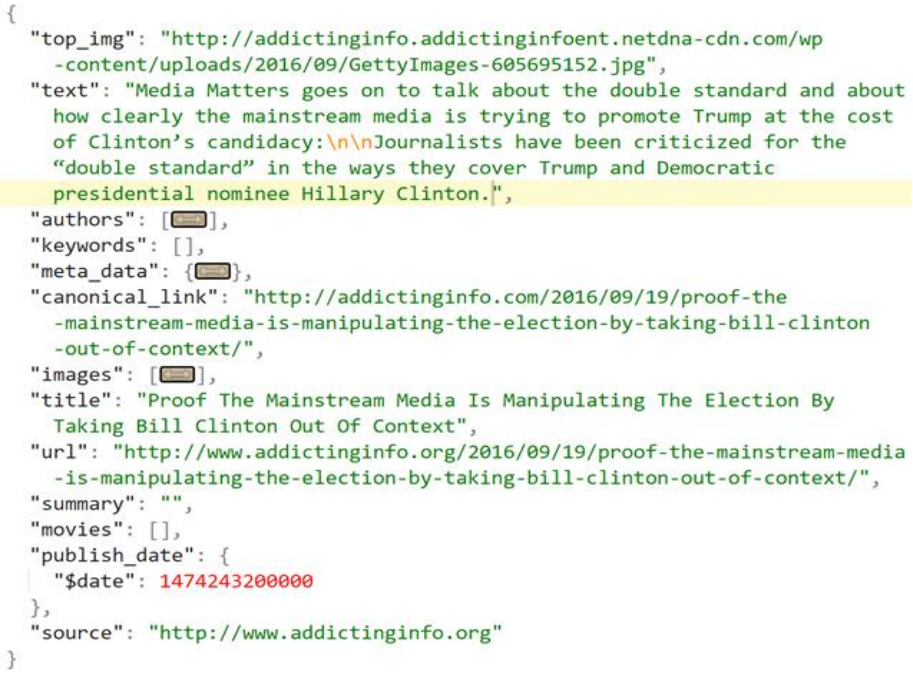
GAT is one such model which utilizes the relationship between news articles and the feature vectors for the classification task. GAT is also a neural network based model. There are other graph-based machine learning neural network models which do the same task but GAT provides attention on certain neural nodes which are considered as important for the task. Other graph-based models treat all nodes in the neural network equally and hence might not provide better accuracy or takes a long time to train. GAT is computationally efficient as well. GAT can parallelize operation on nodes as they need not wait for the result from other nodes. GAT implementation doesn’t include matrix inverse operations which are very costly. These are the advantages of GAT over other machine learning models. Hence in this project, GAT was used as the classification model. The accuracy achieved was around 80% for BuzzFeed dataset and 77% for PolitiFact. The results are very good considering that the credibility of the user is not present in the dataset and not a lot of hyperparameter tuning was done while training the GAT model. With better tuning, even better results can be achieved.

5. DATASETS (Descriptions, Sizes and Preprocessing Steps)

Fake news detection in social media aims to extract useful features and build effective models from existing social media datasets for detecting fake news in the future. Thus, a comprehensive and large-scale dataset with multidimensional information in online fake news ecosystem is important. The multidimensional information not only provides more signals for detecting fake news but can also be used for research such as understanding fake news propagation and fake news intervention. Though there exist several datasets for fake news detection, the majority of them only contain linguistic features. Few of them contain both linguistic and social context features. To facilitate research on fake news, we used a publicly available data repository called FakeNewsNet [3] includes not only news contents and social contents, but also spatiotemporal information. To collect reliable ground truth labels for fake news, fact-checking websites were used to obtain news contents for fake news and true news such as PolitiFact and BuzzFeed news. Buzzfeed and Politifact news is stored in json format and contains the following fields: image URL, text, authors, keywords, metadata, title, URL, publish date and source.

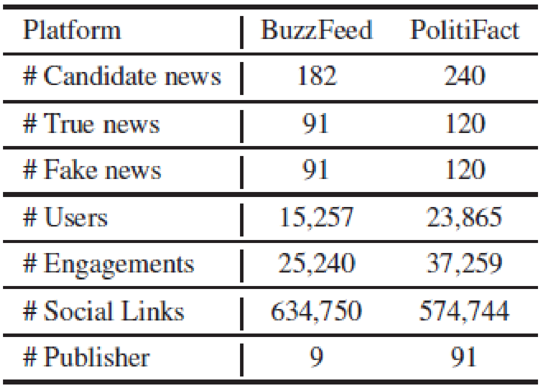
BuzzFeedNews: This dataset comprises of a complete sample of news published on Facebook from 9 news agencies over a week close to the 2016 U.S. election from September 19 to 23 and September 26 and 27. Every post and the linked article were fact-checked claim-by-claim by 5 BuzzFeed journalists. It contains 1,627 articles 826 mainstream, 356 left-wing, and 545 right-wing articles.

Politifact: In PolitiFact, journalists and domain experts review the political news and provide fact-checking evaluation results to claim news articles as fake or real. We utilize these claims as ground truths for fake and real news pieces. In PolitiFact's fact-checking evaluation result, the source URLs of the web page that published the news articles are provided, which can be used to fetch the news content related to the news articles. In some cases, the web pages of source news articles are removed and are no longer available. The figure below represents a sample fake json file.



```
{
  "top_img": "http://addictinginfo.addictinginfoent.netdna-cdn.com/wp
    -content/uploads/2016/09/GettyImages-605695152.jpg",
  "text": "Media Matters goes on to talk about the double standard and about
    how clearly the mainstream media is trying to promote Trump at the cost
    of Clinton’s candidacy:\n\nJournalists have been criticized for the
    “double standard” in the ways they cover Trump and Democratic
    presidential nominee Hillary Clinton.",
  "authors": [⋯],
  "keywords": [],
  "meta_data": {⋯},
  "canonical_link": "http://addictinginfo.com/2016/09/19/proof-the
    -mainstream-media-is-manipulating-the-election-by-taking-bill-clinton
    -out-of-context/",
  "images": [⋯],
  "title": "Proof The Mainstream Media Is Manipulating The Election By
    Taking Bill Clinton Out Of Context",
  "url": "http://www.addictinginfo.org/2016/09/19/proof-the-mainstream-media
    -is-manipulating-the-election-by-taking-bill-clinton-out-of-context/",
  "summary": "",
  "movies": [],
  "publish_date": {
    "$date": 1474243200000
  },
  "source": "http://www.addictinginfo.org"
}
```

Fig 3: Politifact sample file



Platform	BuzzFeed	PolitiFact
# Candidate news	182	240
# True news	91	120
# Fake news	91	120
# Users	15,257	23,865
# Engagements	25,240	37,259
# Social Links	634,750	574,744
# Publisher	9	91

Fig 4: Statistics of Datasets

6. EVALUATIONS (Metrics, Experiments, Findings)

To evaluate the performance of fake news detection algorithms, we used accuracy to evaluate classifiers in related areas:

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|}$$

where TP, FP, TN, FN represent true positive, false positive, true negative and false negative, respectively. We compared the proposed framework to other machine learning algorithms. To get the following results:

Dataset	Bagging	Multi NB	LR	AdaBoost	RF	GB	DT	MLP	SVC	KNN	GA T
Buzzfeed	0.54	0.49	0.51	0.60	0.54	0.65	0.49	0.54	0.59	0.62	0.81
Politifact	0.54	0.41	0.43	0.50	0.44	0.50	0.56	0.46	0.46	0.46	0.76

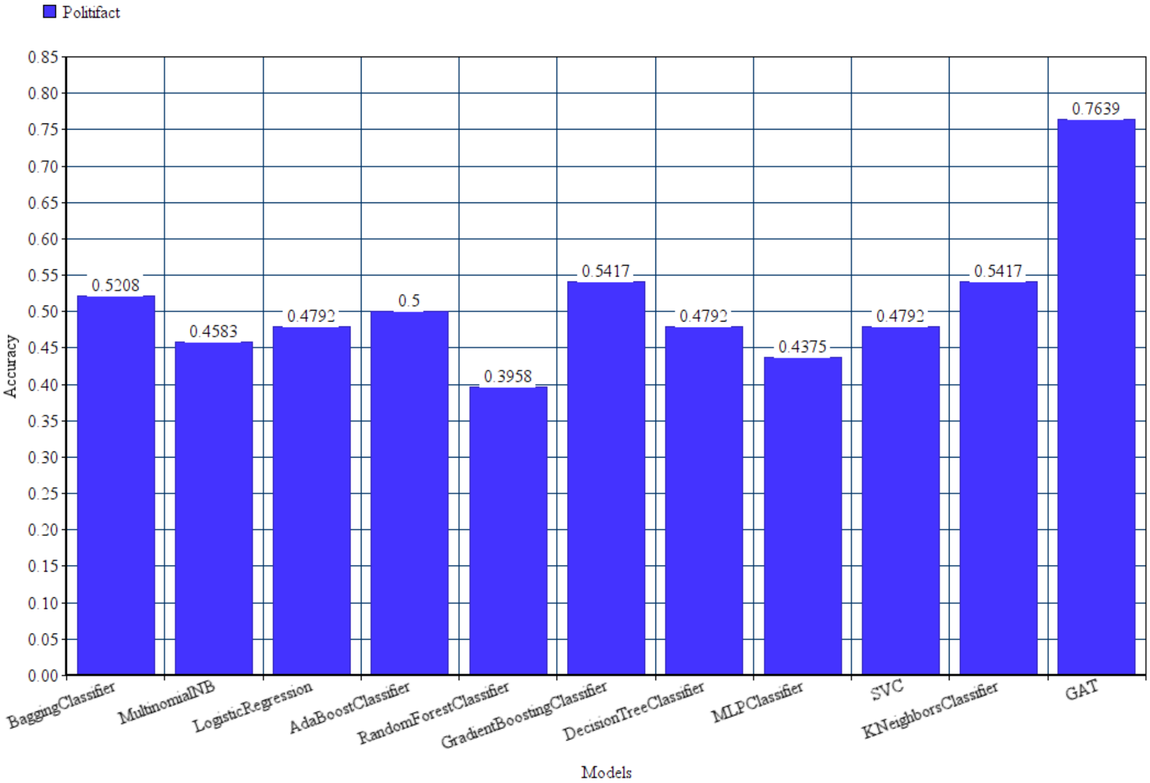


Fig 5: Test Data Result for BuzzFeed

Impact of Training Data Size

We further investigate whether larger amounts of training data can improve the identification of fake news. We plot the learning curves with respect to different training data size, as shown in the below figure. By plotting these learning curves, we can see that the detection performance tends to increase with the increase of training ratio for all compared methods on both datasets.

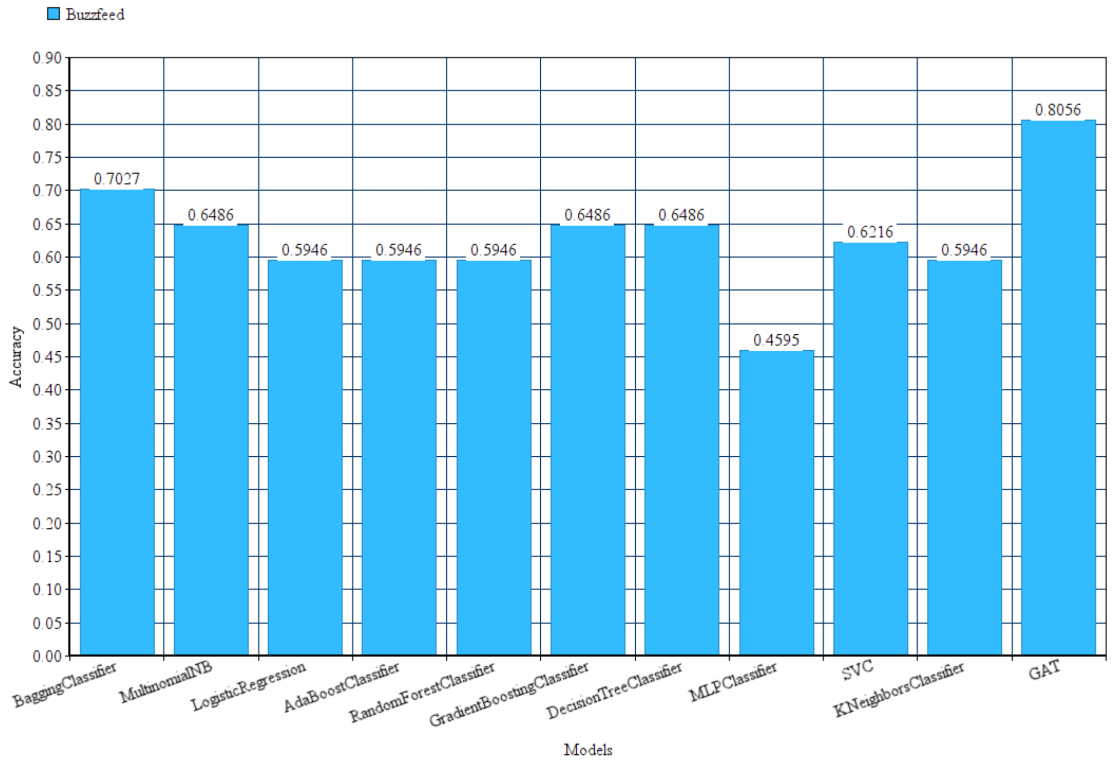


Fig 6: Test Data Result for PolitiFact

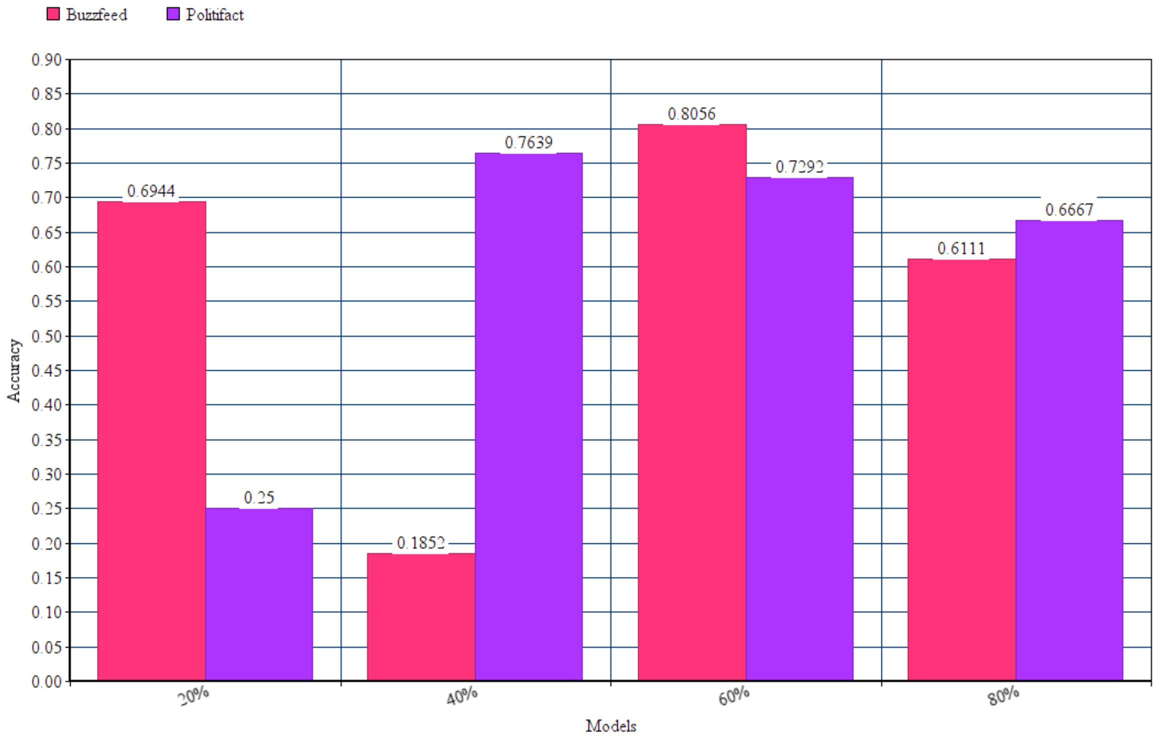


Fig 7: Size of Training Dataset vs Test Accuracy

7. DIVISION OF WORK AND TEAM MEMBERS' CONTRIBUTIONS

Team Members:

a. Abhijith Shreesh (ASU ID: 1213204276)
b. Aditya Chayapathy (ASU ID: 1213050538)
c. Anuhya Sai Nudurupati (ASU ID: 1212931887)
d. Arun Karthick Manickam Alagar Muthumanickam (ASU ID: 1213135077)
e. Jagdeesh Basavaraju (ASU ID: 1213004713)

Tasks involved and division of work:

a. Data Pre-processing - Arun Karthick and Abhijith
b. Feature extraction (BERT) - Abhijith and Aditya
c. Deriving relationships (Adjacency Matrix) - Aditya and Jagdeesh
d. Classification model tuning and training (GAT) - Anuhya and Jagdeesh
e. Performance evaluation of the models trained - Anuhya and Arun Karthick
f. Report and documentation - All members involved

8. CONCLUSION

Within the vast domain of social media and its issues, there are critical problems which are a threat to the social media as a platform. Among those, fake news are not platform specific and more critical because of its effect over the drop in platform usage. This is framework, we are trying to build a proactive methodology which could leverage the logistical information for better classification. This method is not only specific to any platform or social media, but the same knowledge can be applied in the domain for Q&A forums and blogs such as StackOverflow, where the auxiliary users information in available. The future work for this framework involves integration with multiple platforms to know more about the user profile via a combinatorial credibility score. Also, making this score common across all platforms in the interest will have a more positive impact on the interest.

9. REFERENCES

1. Beyond News Contents: The Role of Social Context for Fake News Detection; Kai Shu, Suhang Wang, Huan Liu; 2019 Association for Computing Machinery.
2. Graph Attention Networks; Petar Velickovi, Guillem Cucurul, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio; ICLR 2018.
3. FakeNewsNet: A Data Repository with News Content, Social Context and Spatial temporal Information for Studying Fake News on Social Media; Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee and Huan Liu; AAAI 2019.
4. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova; arXiv preprint arXiv:1810.04805 2018;
5. Fake News Detection on Social Media: A Data Mining Perspective; Shu, Kai and Sliva, Amy and Wang, Suhang and Tang, Jiliang and Liu, Huan; ACM SIGKDD Explorations Newsletter 2017.