

A RELIABLE ARTIFICIAL INTELLIGENCE MODEL FOR FALSE NEWS DETECTION MADE BY PUBLIC FIGURES

Prashant Kumar Shrivastava¹, Mayank Sharma², Megha Kamble³, Vaibhav Gore⁴

^{1,3,4}Computer Science & Engineering, Lakshmi Narain College of Technology
Excellence,

^{1,3,4}LNCTE Bhopal ,

²ITM University ,Gwalior, India

^{1,2,3,4}Computer Science & Engineering

ABSTRACT

The quick access to information on social media networks as well as its exponential rise also made it difficult to distinguish among fake information or real information. The fast dissemination by way of sharing has enhanced its falsification exponentially. It is also important for the credibility of social media networks to avoid the spread of fake information. So it is emerging research challenge to automatically check for misstatement of information through its source, content, or publisher and prevent the unauthenticated sources from spreading rumours. This paper demonstrates an artificial intelligence based approach for the identification of the false statements made by social network entities. Two variants of Deep neural networks are being applied to evaluate datasets and analyse for fake news presence. The implementation setup produced maximum extent 99% classification accuracy, when dataset is tested for binary (true or false) labeling with multiple epochs.

KEYWORDS

fake news; artificial intelligence; deep learning: RNN, RAMP studio.

1. INTRODUCTION

The concept "fake news (FNs)" was much less unheard of and not popular in a few decades, but it emerged as a major monster in this modern age of social media. Fake news, bubbles in information, manipulation of news, or loss of interest in media are growing problems in our culture. However, a clear understanding of FNs as well as its sources is important to begin to resolve this issue. It is only in this context that one should look at the various methods & fields of machine learning (ML), natural language processing (NLP), & artificial intelligence (AI), which could help us face the situation.

The calculation or even the proper definition of fake news could very easily become a subjective issue instead of an objective metric. In its purest nature, fake news is entirely composed, manipulated to imitate trustworthy media, and draws full exposure.

Motivation

The popular issue of fake news is very hard to resolve in today's digital world when thousands of information sharing platforms will distribute FNs or incorrect data. A big issue in AI is that of

developing artificial bots to produce and distribute fake news. [1]. problem is tragic because many people think they read something on the Internet, even those who are amateur or new to modern technologies will easily be tricked. A related issue is a spam or fraudulent emails or communications that may lead to theft. It is also reasonable enough to recognize this problem to solve crime rates, political uneasiness, grief & combat the attempts at distributing false news.

The definition of fake news consists of two parts: authenticity & intent. Authenticity ensures that false news material may be confirmed as such and that conspiracy is not found in fake news and, in most situations, it is impossible to prove true or false. The second part, intent, means that the false information is written to mislead the reader.

There are also tens of millions of languages spoken and each language has its vocabulary, script & syntax. The NLP is an AI branch that involves strategies for using text, creating models, and forecasting. The nature of outline text is of different forms and it is difficult to use simply for different linguistic features or forms, like sarcasm, metaphors, etc. This paper aims to build a framework or model that can utilize details in previous news stories to predict whether or not a news report is false.

1.1. Fake News Types

The authors[2] have provided comprehensive literature and the types of fake news can be summarized as follows, based on literature.

1. **Visual-based:** These fake news items are much more useful as material, including fraudulent images, medical videos, or both [3].
2. **User-based:** Fake news stories produce this form of content, which threatens those age groups, genders, culture, political ideology.
3. **Knowledge-based:** These forms of communications explain some unresolved problems by researchers (so-called) or make users believe it is true.
4. **Style-based:** Journalists who pretend to replicate the style of other accredited journalists are writing blogs.
5. **Stance-based:** It is also a depiction of valid statements so that its meaning and intent are changed.



Figure 1: Different approaches to fake news detection.

Looking at Figure 1, it can be seen that unsupervised & supervised models of learning are primarily based on textual news content. Note that ML models are usually a tradeoff of precision or recall so that a model that is very effective in detected counterfeit news will have a high false-positive rate relative to a model that does not likely show a low false-positive rate. However automated censorship related topics can not be discussed here due to ethical issues.

2. LITERATURE REVIEW

Various researchers have tried several ways to overcome the problem, to test which approach works, and to achieve desired outcomes. A variety of studies have discussed fake news identification methods, including role extraction or model construction, from a data mining viewpoint. Feature extraction approaches (both features of news material and features of the social context) combined with metric assessment using accuracy, reminder, or F1 ratings have been shown to bring informed results. Other parameters such as bot-spam, click-bait, and news source also impact the forecast [4].

These were approached by data mining & NLP, but researchers have become increasingly involved in heavy NN-oriented methods with more & more research & development in AI. A story illustrates a method for "capturing," "picking" or "integrating." A model of RNNs is created for the identification of fake data. They also used the RNN to compile the temporal pattern of user behaviors around a single article/text. All of this material is used and incorporated into a false news classification model by the authors in [5].

Another paper [6] deals with LSTM or CNN linguistically-infused NN for classifying Twitter messages. The GloVe Library of Pretrained Vectors was used to introduce the linguistic component [6].

This review provides a model for recognizing forged twitter news in the light of automatic forged Twitter news detection, through finding out how to anticipate reliable estimates We have later conducted a separate comparison of the efficiency of classification performance of dataset among five well-known ML algos like SVM, NB Process, LR & RNN. Our experimental findings have shown that the SVM & NB classification is above the other algos. [7].

To solve this issue, this paper explores the approach of NLP & ML. Using bag-of-words, n-grams, count vectorizers, TF-IDF, or data from five classifiers have been equipped to investigate

which of these data sets of labeled news statements works well. The accuracy, the reminder or f1 scores allow us to find the most effective model. [8].

The authors have described in [9] the language cue approaches with ML, word bags, rhetorical structure or discourse analysis, network analysis approaches, or SVM classifiers. These text-based models do not or do not boost existing approaches.

In [10], authors appreciated or analyzed the principles, methods, or algos used for categorizing fake or fabricated social network news items, authors & subjects. the paper highlighted the study challenge by exposed features of fake news & numerous similarities between news articles, writers & topics. authors of the paper address the FakeDetector model of auto-found news. It creates a deep, pervasive network model based on text and helps the author and subject to learn about the presentation of news articles simultaneously.

3. RESEARCH METHODOLOGY

Problem Formulation

Social media is a double-edged sword for news consumption. On the one hand, the low cost, quick access & fast circulation lead users in the social media to browse and receive content. It makes, on the other hand, a wide variety of " FNs " with purposely misleading facts, i.e. bad quality news. The widespread of fake news is very harmful to individuals & society. Consequently, fake social media news detection has recently grown into an emerging study that attracts considerable interest. Fake social media news detection presents unique features or challenges which make existing news algos ineffective or not applicable. Firstly, FNs is written purposely to mislead readers with FNs, which makes the detection of that too difficult & inexplicably dependent on news content; therefore, supplemental data, like user social commitments on social media, must be included in our evaluation. The second challenge, as social commitments of users with FNs, produces big, incomplete, unstructured, and noisy data is to manipulate this auxiliary material.

Data Pre-Processing

It should be pre-processed before the use of AI algos on the results. First of all, only statements themselves were decided for purposes of classification. That means none of the provided metadata is used for grading. This metadata could improve classification algo in the future. The following measures have been used for pre-processing:

- Split the statements into individual tokens (words).
- All numbers are deleted.
- Removal of all marks of punctuation.
- Delete all other characters, not alpha.
- The rest of the tokens should apply a steaming process. In linguistic morphology, the approach used to avoid (or lemmatize) recruitment information is to reduce inflected or derived terms in the terms stem, base, or root shape – typically in writing. This encourages them to use words that are identical to the same ones (for example "write").
- Stop word removal. Word stops are included in basically a kind of text. These terms are popular and do not impact their importance, so it is beneficial to get rid of them [6].
- Substitute terms with tf-idf values. In information collection tf – idf is a figural metric that represents the meaning of a word for a record in a collection or corpus and is a short term for frequency-inverse document frequency. [7].

1. Accuracy classification based on six available categories
2. Accuracy in binary classification: This metric is as if 2 possible categories were used in the statement – true (according to the last three categories described above) or false (based on the first three above mentioned categories).

Proposed Approach

Several AI algos were used for statement classification during the current study, of which DNN provided the best results. With some parameter tuning, this work selected the modified NN as the approach proposed that is superior to DNN. This is implemented with available standard libraries of a scikit-learn (a programming language library of Python). Two separate measures have been evaluated for all algos:

- Accuracy classification based on 6 obtainable category.
- Binary accuracy of the classification: This metric is regarded as being only available in 2 possible categories, real (based on the last three categories mentioned above).

Recurrent Neural Network

RNN is NN that is specialized in data processing $x(t) = x(1), \dots, x(\tau)$ with an index time step of t between 1 and μ . It is also easier to use RNNs for tasks containing sequential inputs like speech and language. RNN is called repeating since it achieves the same purpose for each part of a sequence, based on the previous calculations.

Architecture: Let's go over an important RNN network briefly.

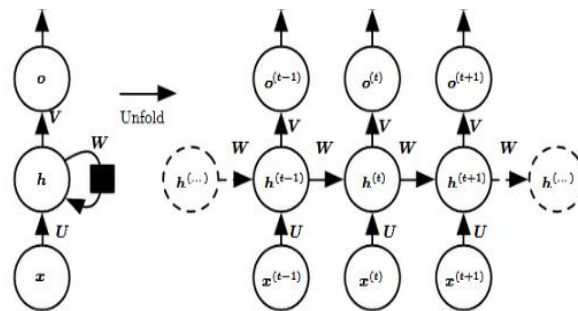


Fig 2: Architecture of RNN

The left side of the diagram showed an RNN notation as well as an RNN was unrolled (or unfolded) into a whole network on the right-hand side. We say we write down the system for the whole sequence. by unrolling. If the sequence we care about is a sentence of three terms, for example, the network will be deployed into a network of three layers, one layer for each word.

Input: $x(t)$ shall be taken as network input at step t .

Hidden state: $h(t)$ is hidden state at time t , function as network memory. $h(t)$ is present input or hidden state for the previous time stage: $\mathbf{h}(t) = f(U \mathbf{x}(t) + W \mathbf{h}(t-1))$.

Weights: The RNN has contributions to hidden connections, a U weight matrix parameterized, the W weight matrix parameterized, or V weight matrix parameterized, to hidden recurring connections, which are distributed over time (U, V, W).

Output: $o(t)$ illustrates the output of the network.

4. RESULTS ILLUSTRATIONS

A RAMP studio, a famous machine learning workflow research team [3] has compiled the data set used for training & testing. It includes brief statements by revered persons available on social media. For statement 6 primary labels were available. The same dataset is taken from web repository and binary and multi class deep neural networks are applied with parameter tuning for optimized results.

```
accuracy = accuracy_score(y_train, pred)
precision = precision_score(y_train, pred, average='micro')
recall = recall_score(y_train, pred, average='micro')
f1 = f1_score(y_train, pred, average='micro')

print('Accuracy :', accuracy*100, '%')
print('Recall :', recall*100, '%')
print('Precision :', precision*100, '%')
print('F1 Score :', f1*100, '%')

Accuracy : 86.08799048751486 %
Recall : 86.08799048751486 %
Precision : 86.08799048751486 %
F1 Score : 86.08799048751486 %
```

Fig 3: Binary DNN

```
pred=y_pred.copy()
pred[pred>=thresholds]-1
pred[pred<thresholds]=0
accuracy = accuracy_score(y_train, pred)
precision = precision_score(y_train, pred, average='micro')
recall = recall_score(y_train, pred, average='micro')
f1 = f1_score(y_train, pred, average='micro')

print('Accuracy :', accuracy*100, '%')
print('Recall :', recall*100, '%')
print('Precision :', precision*100, '%')
print('F1 Score :', f1*100, '%')

Accuracy : 83.7230808824152 %
Recall : 83.7230808824152 %
Precision : 99.93691846711876 %
F1 Score : 91.11430625449317 %
```

Fig 4: Multi DNN

Figures 3 & 4 show the results of the simulation tool showing multiple parameters of accuracy in the case of DNN of two different classes.

```
pred=y_pred.copy()
pred[pred>=thresholds]-1
pred[pred<thresholds]=0
accuracy = accuracy_score(y_train, pred)
precision = precision_score(y_train, pred, average='micro')
recall = recall_score(y_train, pred, average='micro')
f1 = f1_score(y_train, pred, average='micro')

print('Accuracy :', accuracy*100, '%')
print('Recall :', recall*100, '%')
print('Precision :', precision*100, '%')
print('F1 Score :', f1*100, '%')

Accuracy : 99.73576430175717 %
Recall : 99.73576430175717 %
Precision : 99.73576430175717 %
F1 Score : 99.73576430175717 %
```

Fig 5: Binary RNN

```

pred[pred>thresholds]=1
pred[pred<thresholds]=0
accuracy = accuracy_score(y_train, pred)
precision = precision_score(y_train, pred, average='micro')
recall = recall_score(y_train, pred, average='micro')
f1 = f1_score(y_train, pred, average='micro')

print('Accuracy :',accuracy*100,'%')
print('Recall :', recall*100,'%')
print('Precision :',precision*100,'%')
print('F1 Score :',f1*100,'%')

Accuracy : 86.88069758224336 %
Recall : 86.8939093671555 %
Precision : 99.96960024319806 %
F1 Score : 92.97427198190556 %
    
```

Fig 6: Multiclass RNN

Figures 5 and 6 visualize the result of RNN in both binary and multi-class analysis. From all the figures it has been noticed that RNN has given the best accuracy in all the parameters of accurate measurements.

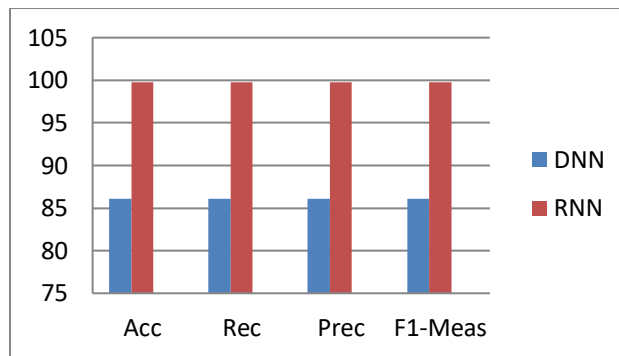


Fig 7: Comparison graph of Binary class of DNN and RNN

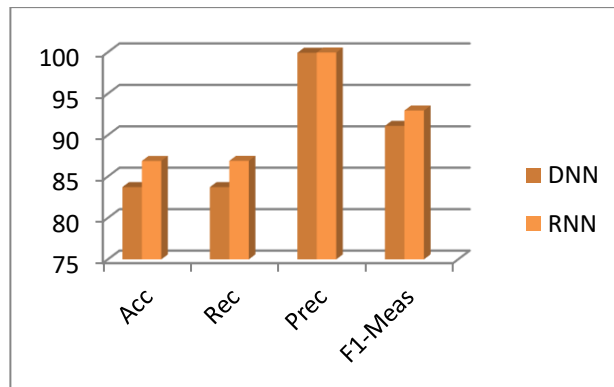


Fig 8: Comparison graph of multi-class of DNN and RNN

In figures 7 & 8, the outcomes of the summarized classifications are shown. The present NN provides the best outcomes in the accuracy of both multiclass and binary classes.

5. CONCLUSION

The fake news challenge is risky or quickly spreads like a wildfire since information can be readily reached in different ways. In this paper, we have compared two NN models for checking the verification of data extracted from the RAMP studio. From the results predicted it clearly

shows how RNN has given the best accuracy on the same dataset as was used in the previous work. Although fake news or messages using different ML methods can be identified successfully in many previous papers. Deep neural network works better in non linear voluminous data set. Hence RNN is found to be the best method in the classification of the news. In the future, we need to improve the results of Multiclass DNN and RNN as RNN can give an accuracy of 86% in a multiclass dataset. A new technique is hence needed to invent for this purpose.

REFERENCES

- [1] Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96-104.
- [2] Hunt, E. (2016). What is fake news? How to spot it and what you can do to stop it. *The Guardian*, 17
- [3] Golbeck, Jennifer, et al. "A Large Labeled Corpus for Online Harassment Research" Proceeding of the 2017 ACM on Web Science Conference. ACM, 2017
- [4] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
- [5] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). CSI: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806). ACM.
- [6] Volkova, S., Shaffer, K., Jang, J. Y., & Hodas, N. (2017, July). Separating facts from fiction: Linguistic models to classify suspiciously and trusted news posts on twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 647-653).
- [7] Abdullah-All-Tanvir, Mahir, E. M., Akhter, S., & Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 2019 7th International Conference on Smart Computing & Communications (ICSCC).
- [8] Vasu Agarwala, H. Parveen Sultanaa, Srijan Malhotra, Amitrajit Sarkar "Analysis of Classifiers for Fake News Detection" INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019
- [9] Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community* (p. 82). American Society for Information Science.
- [10] Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with the deep diffusive network model. arXiv preprint arXiv:1805.08751.