International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3561–3571

Fake News Detection in Social Media Using Deep Neural Networks

Dr SusmithaValli Gogula[1], Dr B.Madhuravani[2], L.Alekya[3]

[1,2,3]*Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad.*

Abstract

In the contemporary era, there has been increased usage of social networks through which news are spreading faster. They became virtual platforms for instant communications. Even governments and organizations are using social networks to convey intended news to people. It is good to have such media in order to improve communication across communities or people. However, there are some incidents where it is proved that fake news is spread by some people intentionally. Such news is created to damage impression of an organization or person or agency. This became a tool to damage products and services of opponents in case of businesses. Dissemination of fake news has its consequences and there is need for preventing it in social media. In literature there are many methods found to prevent spreading of fabricated news. However, in the wake of computation innovations and deeplearning methods, it is possible to improve the state of the art. Towards this end, in this paper, we proposed a framework for fake news detection using Convolutional Neural Network (CNN) which is one of the deep learning methods. A prototype application is built using Python data science platform to show the utility of the proposed system. The empirical results revealed that the proposed algorithm showed better performance over its predecessors.

Keywords –Classification, Deep learning, Fake news detection, Social media, Supervised learning

1. INTRODUCTION

Social media such as Twitter and Facebook became very important part of lives of people of all walks of life. Their usage is drastically increasing. There is provision for instant messaging and opinion sharing. Social media has also become a platform for sharing news. In this context it is observed that there is practice of fake news or fabricated news to favour particular party or company or individual and cause damage to some other party [1]. Therefore, it is essential to have methods to prevent fake news from being spread by individuals who have malicious intentions. It may be spread through social media, mails and even advertisements. There are malicious advertisements that are to be identified. As the news over social media have high impact on people, it is essential to curb fake news by developing efficient methods.

From the literature, many fake news detection methods are found. Different types of fake news are identified in [2]. Automatic detection of fake news is the main focus in [3] and [4]. Identification of phishing mails and malware is the objective of the research carried out in [5] while the procedure to stamp out fake news is discussed in [6]. There are deep learning methods found in the literature. In [7] deep learning based supervised learning is explored for fake news detection effectively. In [15] also deep learning is used for automatic identification of fake news. From the review of literature, it is understood that there is need for further research on fake news detection using deep learning methods. Our contributions in this paper are as follows.

1. Proposed methodology based on deep learning for fake news detection over social media.
2. Proposed two deep learningmodels known as CNN baseline model and CNN advanced model[18].
3. A prototype application is developed using Python data science platform for implementation of the models using Keras and TensorFlow.

The remainder of the paper is structured as follows. Section 2 reviews literature on existing fake news detection models. Section 3 presents the proposed solution along with a framework and deep learning

ISSN: 2233-7857 IJFGCN

models. Section 4 provides details of experimental setup. Section 5 presents experimental results while Section 6 concludes the paper and gives directions for future scope of the research.

2. LITERATURE SURVEY

This section reviews literature on fake review detection methods existing. Westerman*et al*. [1] opined that social media became an important information source. They investigated on the information credibility in the content of social media and found that fake news or manipulated news cause significant damage to persons or organizations. Rubin *et al*. [2] did research in similar lines as in [1] and found that deceptive news exists over social media and one needs to be careful about it. They found three kinds of fake news over social media. Chen *et al*. [3] focused on the automatic detection of fake news over social media. They proposed a methodology and also established the need for detection of fake news. Marsi*et al*. [4] investigated on malicious advertisements over social media. They developed a method to find the advertisements that are not genuine. Qbeitah and Aldwairi [5] studied emails that are associated with phishing and malicious software or malware.

When fake news is detected, there needs to be a mechanism that will stamp fake news as such according to Pogue [6]. Konagala and Bano [7] employed deep learning for fake news detection. It is based on semantic similarity and supervised learning over social media data. Balmas [8] observed that there is enough evidence of fake news over social media. It is associated with multiple news sources associated with cynicism, alienation and political attitudes. Aldwairi *et al*. [9] focused on the machine learning based solution to find drive-by download attacks. Aldwairi*et al*. [10] on the other hand focused on malicious URLs that cause inconvenience and damage to information systems using supervised learning. Messabi*et al*. [11] proposed a method based on domain name features and DNS records in order to detect spreading of malware.

Abu-Nimeh*et al*. [12] investigated on the spam posts and malicious posts over social media. They found that it is possible to detect such posts and eliminate them. Brewer *et al*. [13] studied the impact of real news in the presence of fake news. They observed that fake news is capable of damaging information systems to a greater extent. Monti *et al*. [14] proposed a methodology based on deep learning for fake news detection. It is known as Geometric deep learning. Qawasmeh*et al*. [15] also employed deep learning for detection of fake news automatically. From the review of literature, it is understood that there is need for further research on fake news detection using deep learning methods.

3. METODOLOGY

We proposed a framework for the research on the fake news detection. First, we proposed a CNN baseline model for empirical study on fake news dataset collected from Kaggle [16]. Afterwards, the CNN advanced model is used in order to have more efficient detection of fake news. There is a traditional method known as Naïve Bayes also which is used to compare the performance difference between conventional and deep learning models. The proposed framework is based on supervised learning. Figure 1 shows the framework with the procedure illustrated for fake news detection.

ISSN: 2233-7857 IJFGCN

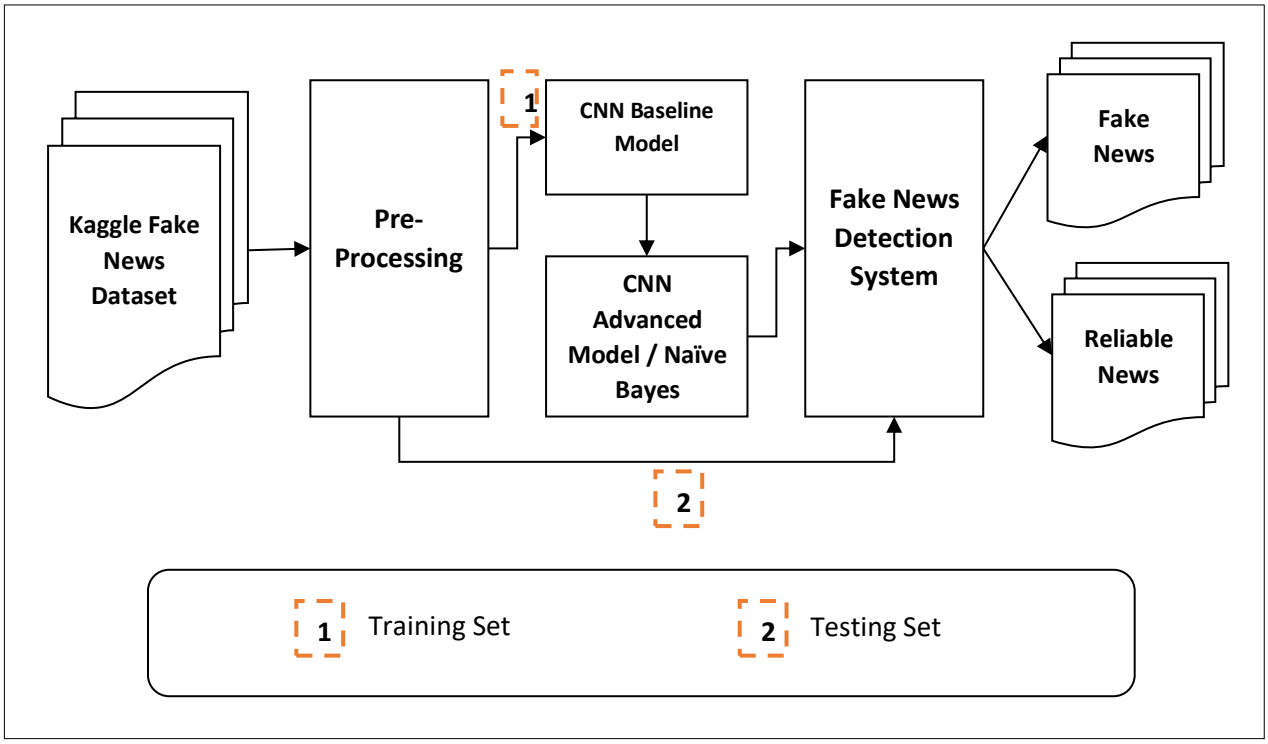
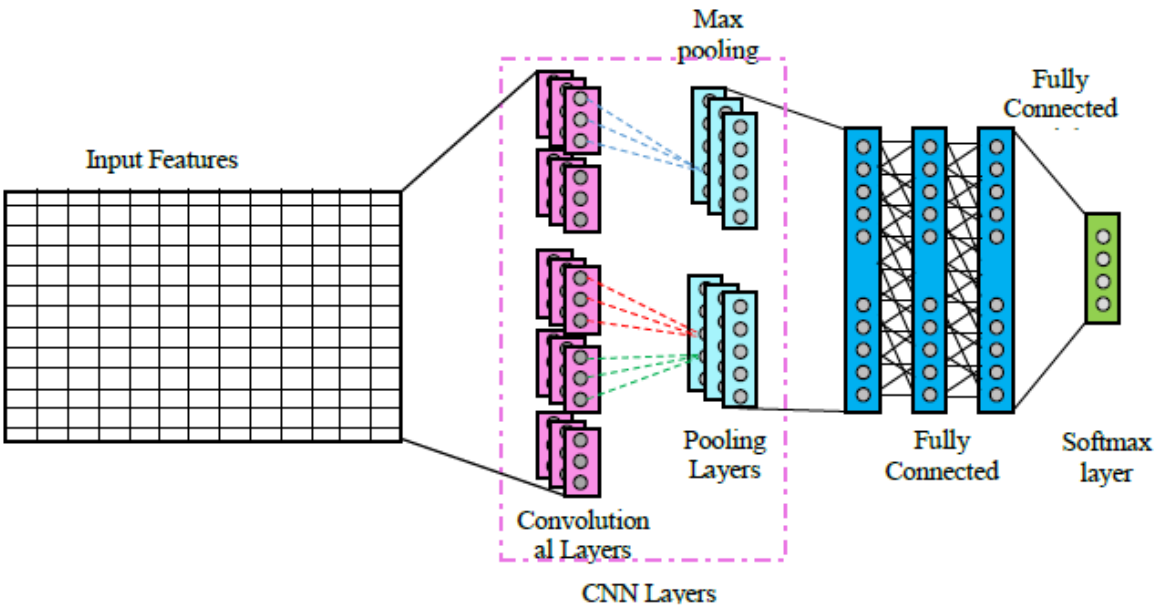


Figure 1: Overview of the proposed framework for fake news detection

As presented in Figure 1, the Kaggle fake news dataset is taken and itis subjected to pre-processing to generate both training and testing sets. The training set is given to the CNN baseline model, CNN advanced model and the Naïve Bayes classification algorithm[18]. The result of these models is the fake news detection system learned from given deep learning or machine learning method (supervised learning). The detection system then classifies the testing set into fake or reliable news. First, a Convolutional Neural Network (CNN) is built using Keras of Python data science platform. TensorFlow is used as backend. This CNN model is known as baseline model without optimizations using convolutions. Afterwards, the CNN model is used with advanced convolutions. This model is known as the CNN advanced model. CNN is an improved[19] neural network (NN) model containing many layers of learning. Therefore, it is known as deep learning. With deep learning, the learning process assumes depth in learning process.



ISSN: 2233-7857 IJFGCN

Figure 2: Deep Convolutional Neural Network for fake news detection

As presented in Figure 2, the fake news dataset is taken from Kaggle and its features are given as input to the deep learning method. Two important layers are found in the deep learning architecture. They are called as convolutional layers and pooling layers. The former is meant for filtering process and the latter is used for performing efficient subsampling. The feature filtering is used to reduce number of dimensions that leads to quality improvement. And the subsampling process is meant for enhancing quality in training process. While learning, each layer predicts weights associated with features and the weights may be adjusted as the learning process is in progress. Finally, it selects the features that help in predicting class labels accurately. There are many iterations in the process of learning. Each iteration is known as an epoch. After training fake review prediction model, it gets discriminative power that will be better than the traditional supervised learning algorithms used for classification. After getting maxpooling results, they are given to the fully connected layer. This layer is the final layer in the prediction process which results in the identification of fake news and discriminating them from reliable news.

4. EXPERIMENTAL SETUP

Experiments are made with the prototype application built using Python data science platform. Anaconda for Python version 3.7 is used for empirical study. TensorFlow is used as backend. Natural Language Processing (NLP) packages, keras, numpy and pandas packages are used in the implementation of the proposed system.

	A	B	C	D	E
1	id	title	author	text	label
2	0	House Dem Aide: We Didnâ€™t Even See Comeyâ€™s Letter U	Darrell Lucus	House Dem Aide: We Didnâ€™t Even See	1
3	1	FLYNN: Hillary Clinton, Big Woman on Campus - Breitbart	Daniel J. Flynn	Ever get the feeling your life circles the roundabou	0
4	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29,	1
5	3	15 Civilians Killed In Single US Airstrike Have Been Identified	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstrike	1
6	4	Iranian woman jailed for fictional unpublished story about wo	Howard Portnoy	Print	1
7	5	Jackie Mason: Hollywood Would Love Trump if He Bombed No	Daniel Nussbaum	In these trying times, Jackie Mason is the Voice of	0
8	6	Life: Life Of Luxury: Elton Johnâ€™s 6 Favorite Shark Pictures T	nan	Ever wonder how Britainâ€™s most iconic pop	1
9	7	BenoÃ®t Hamon Wins French Socialist Partyâ€™s Presidential	Alissa J. Rubin	PARIS â€" France chose an idealistic, traditional	0
10	8	Excerpts From a Draft Script for Donald Trumpâ€™s Q&ampA V	nan	Donald J. Trump is scheduled to make a highly anti	0
11	9	A Back-Channel Plan for Ukraine and Russia, Courtesy of Trum	Megan Twohey and Sc	A week before Michael T. Flynn resigned as nation	0
12	10	Obamaâ€™s Organizing for Action Partners with Soros-Linked	Aaron Klein	Organizing for Action, the activist group that morp	0

Figure 3: An excerpt from fake news data set (training set)

Fake news dataset is collected from Kaggle datasets [16]. As presented in Figure 3, the training data has different attributes such as id, title, author, text and label. The id attribute contains a unique identification number for the news article. The title attribute has a title for given news article. The author attribute indicates the writer of the news. The text attribute holds actual news content[21]. The label attribute on the other hand contains class label with value 1 indicating unreliable (fake news) and 0 indicates reliable (genuine news).

5. EXPERIMENTAL RESULTS

Fake review detection models are made as per the proposed methodology. Two models are proposed based on Convolutional Neural Network (CNN), a deep learning method. After building the two models, the results are compared with a traditional machine learning model known as Naïve Bayes. All models are supervised learning models that are trained prior to detecting[17] fake news.

5.1 Results of CNN Baseline Model

ISSN: 2233-7857 IJFGCN

CNN baseline model results are presented in this section. It includes the results pertaining to model accuracy, model loss and confusion matrix. Model accuracy is measured for both training and test sets. Figure 3 visualizes the model accuracy of the CNN baseline model.

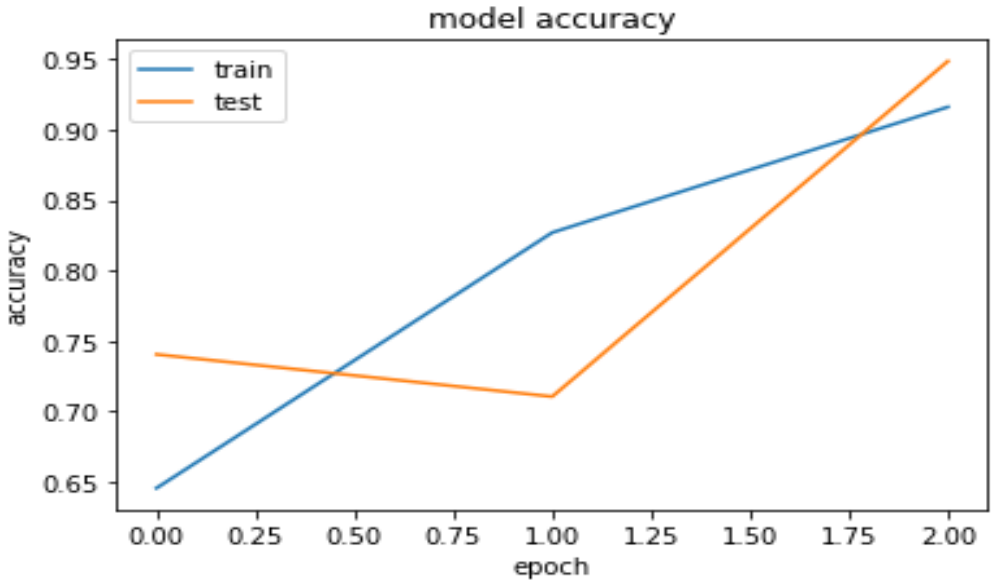


Figure 4: Model accuracy of the CNN baseline model

As presented in Figure 4, the epoch values are presented in horizontal axis and vertical axis shows the accuracy of the model for both training and testing sets. The results reveal that the epoch value influences the model accuracy. As the epoch value is increased the accuracy is also increased in case of training set. In case of test set, it is true but initially it is decreased till epoch 1.00 and then increased gradually. Overall, the training set showed higher level of accuracy.

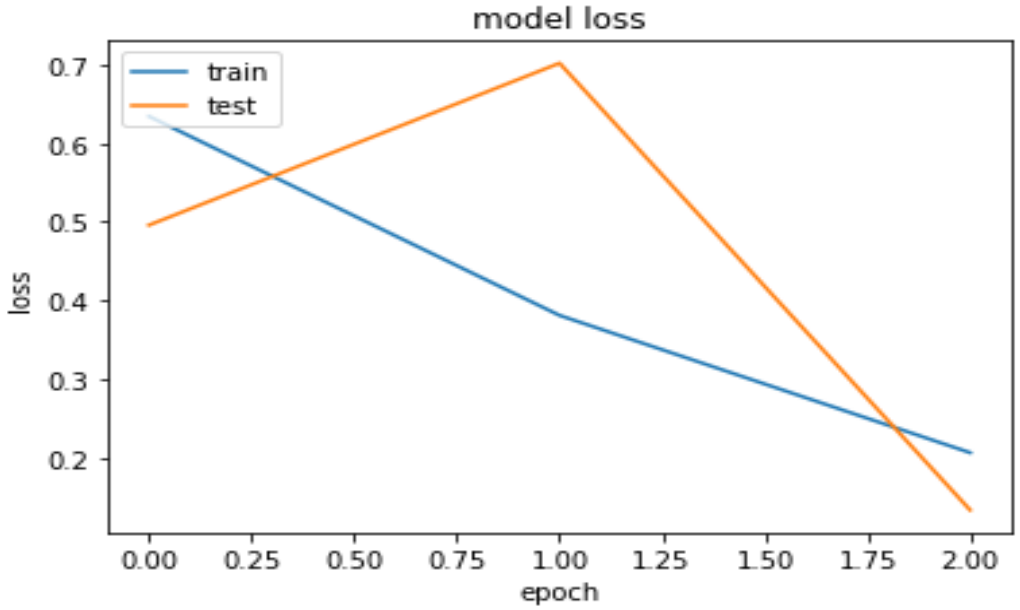


Figure 5: Model loss of the CNN baseline model

As presented in Figure 5, the epoch values are presented in horizontal axis and vertical axis shows the loss of the model for both training and testing sets. The results reveal that the epoch value influences the model loss. As the epoch value is increased the model loss is decreased in case of training set. In case of test set, it is true but initially it is increased till epoch 1.00 and thendecreased gradually. Overall, the training set showed higher level of accuracy.

ISSN: 2233-7857 IJFGCN

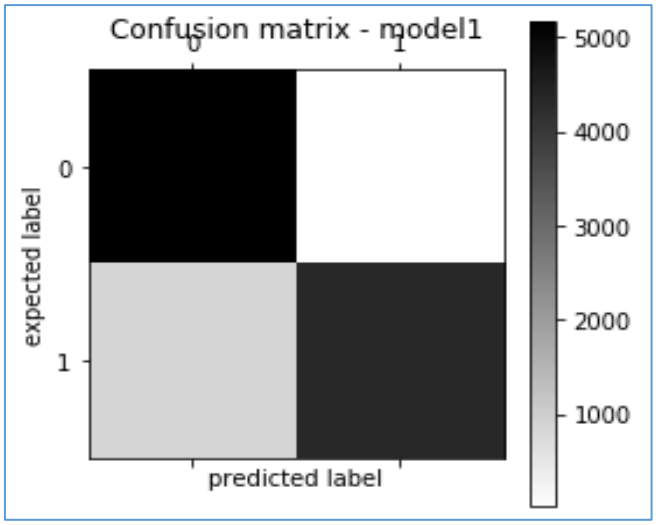


Figure 6: Confusion matrix of CNN base model

As presented in Figure 6, the confusion matrix shows the true positives, false positives, true negatives and false negatives. They are used in the computation of model accuracy. Both expected and predicted labels are provided and based on the colour values, it is possible to ascertain the values for the true positive and other measures.

5.2 Results of CNN Advanced Model

This subsection provides results of CNN advanced model for both training and testing sets. The results are provided in terms of model accuracy and model loss.

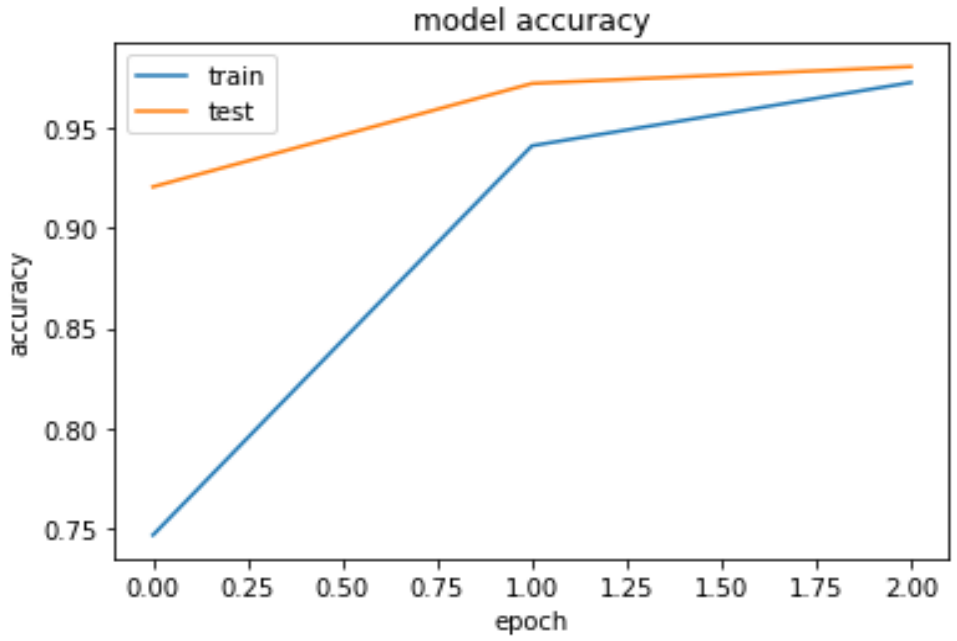


Figure 7: Model accuracy of the CNN advanced model

As presented in Figure 7, the epoch values are presented in horizontal axis and vertical axis shows the accuracy of the model for both training and testing sets. The results reveal that the epoch value influences the model accuracy. As the epoch value is increased the accuracy is also increased in case of training set and testing set. Overall, the testing set showed higher level of accuracy. When compared with the CNN baseline model, the CNN advanced model showed higher level of accuracy.

ISSN: 2233-7857 IJFGCN

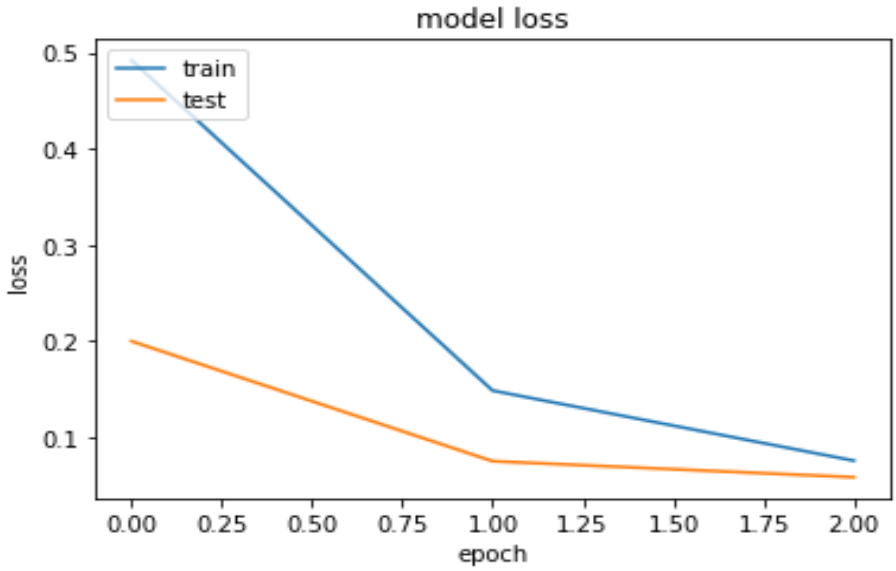


Figure 8: Model loss of the CNN advanced model

As presented in Figure 8, the epoch values are presented in horizontal axis and vertical axis shows the loss of the model for both training and testing sets. The results reveal that the epoch value influences the model loss. As the epoch value is increased the model loss is decreased in case of training set and testing set. Overall, the test set showed low level of model loss. However, when compared with the CNN baseline model, the CNN advanced model showed higher level of performance in terms of model loss.

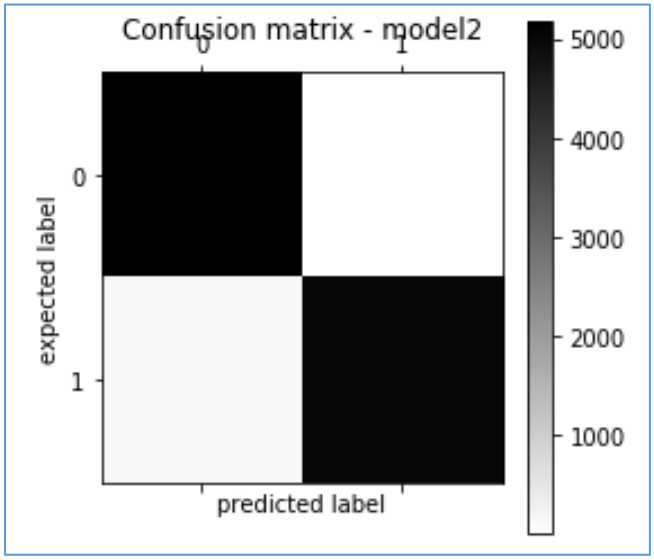


Figure 9: Confusion matrix of CNN advanced model

As presented in Figure 9, the confusion matrix shows the true positives, false positives, true negatives and false negatives. They are used in the computation of model accuracy. Both expected and predicted labels are provided and based on the colour values, it is possible to ascertain the values for the true positive and other measures.

5.3 Performance Evaluation

This sub section evaluates the performance of the proposed deep learning models and compare them with that of traditional machine learning model known as Naïve Bayes.

ISSN: 2233-7857 IJFGCN

Deep Learning Models	No. of Test Samples	Correct Predictions	Model Accuracy
CNN Baseline	10400	9516	0.9150
CNN Advanced	10400	10229	0.9835

Table 1: Shows performance comparison between the proposed deep learning models

As presented in Table 1, the performance comparison is made between CNN baseline model and CNN advanced model. The results are presented in terms of number of test samples, number of correct predictions and model accuracy.

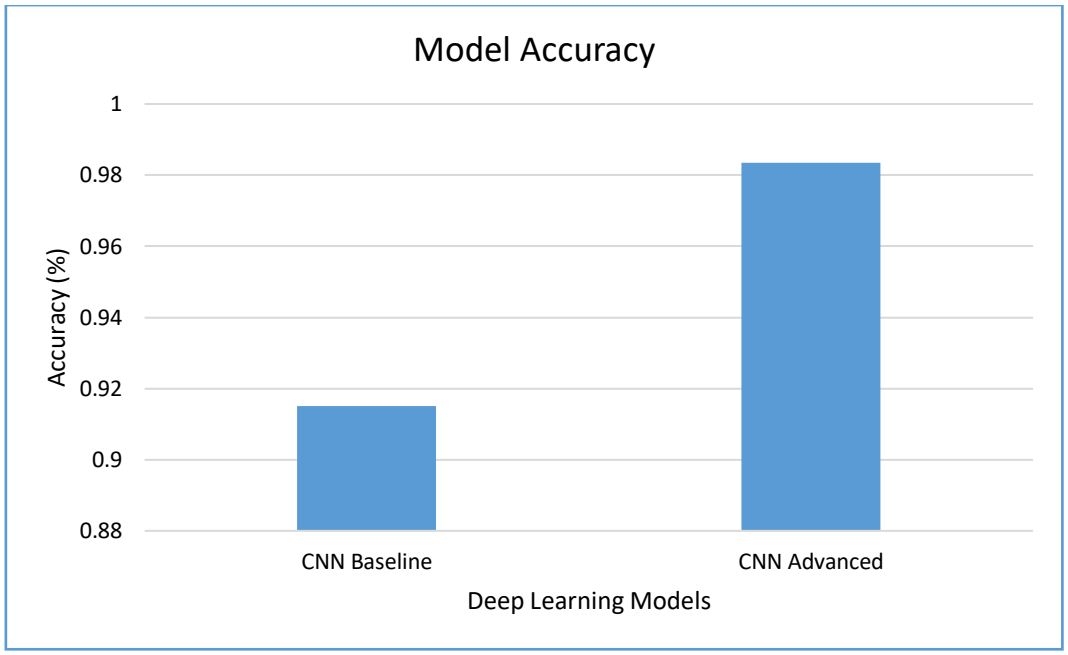


Figure 10: Performance comparison between CNN models

As presented in Figure 10, the deep learning models are shown in horizontal axis while vertical axis shows model accuracy. The results revealed that the CNN advanced model showed better accuracy with 0.9835 which is compared with that of CNN baseline model that shows accuracy as 0.9150. Both models used same number of samples as input but the number of correct predictions differ. CNN baseline made 9516 correct predictions while the CNN advanced model made 10229 correct predictions. This reveals the performance improvement with CNN advanced model.

Deep Learning Models	Model Accuracy (%)
Naïve Bayes	0.8997
CNN Baseline	0.9150
CNN Advanced	0.9835

Table 2: Performance comparison among models

As presented in Table 2, the deep learningmodels are compared with a traditional machine learning model known as Naïve Bayes. The model accuracy is presented for CNN baseline model, CNN advanced model and Naïve Bayes model.

ISSN: 2233-7857 IJFGCN

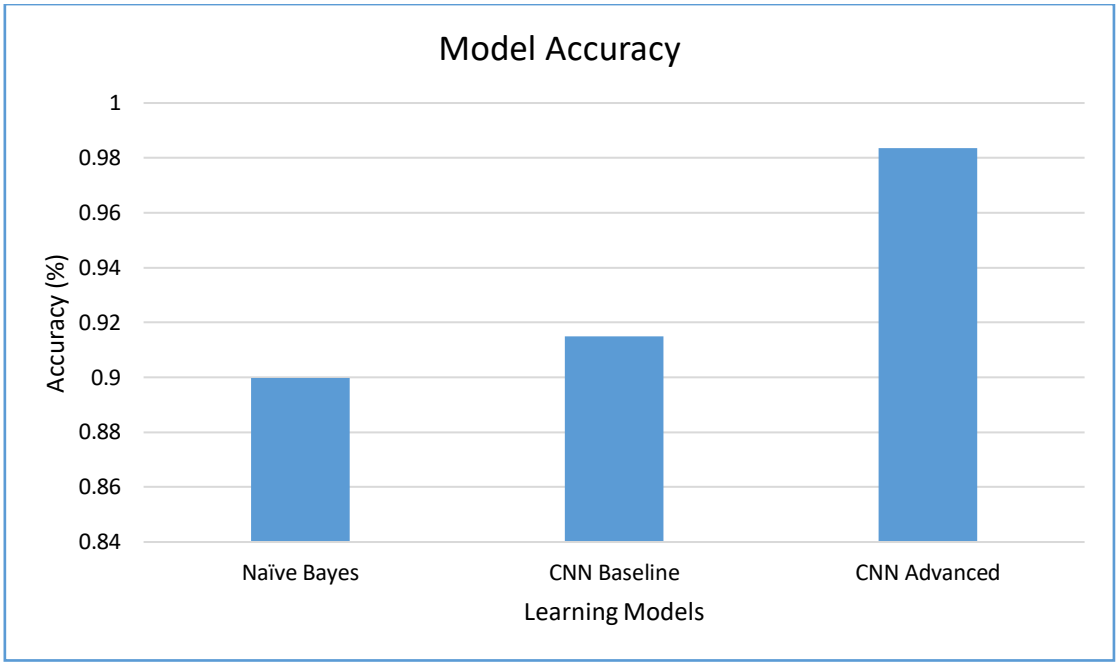


Figure 11: Performance comparison among all models

As presented in Figure 11, the learning models are shown in horizontal axis while vertical axis shows model accuracy. The results revealed that the CNN advanced model showed better accuracy with 0.9835 which is compared with that of CNN baseline model that shows accuracy as 0.9150 and Naïve Bayes model that shows accuracy as 0.8997. All models used same number of samples but the number of correct predictions differed. Both deep learning models showed better performance over the traditional machine learning models.

6. CONCLUSION AND FUTURE WORK

Looking at the current and past scenarios we can deliberately say that the majority of the fake news creators and readers have been tremendously increased. And based on the situations we cannot criticize common people. As living in the busy world every individual doesn't find any difference in real and fake news,in coming days there will be more value for fake news and less value for real news. As of now we are aware of fake news in some or the other instances. Only less amount of people knows what is the truth behind the fabricated news whether it may be real or not. A person creates a piece of false news in social media that piece of news gets widely circulated to about 90% of the social media users and the remaining 10% people may not believe it with having proper details and investigating. To detect whether the news is a real news or fake news we proposed a framework for fake news detection using Convolutional Neural Network (CNN) which is one of the deep learning methods. The proposed CNN advanced model showed highest performance with 0.9835 accuracy. The CNN base model showed 0.915 accuracy while the traditional machine learning method known as Naïve Bayes showed 0.8997 accuracy. The CNN baseline model is better than the Naïve Bayes model[20]. A prototype application is built using Python data science platform to show the utility of the proposed system. The results showed that deep learning-based methods are better than the traditional machine learning methods. In future we intend to improve the deep learning models with techniques like transfer learning.

References

[1] Westerman D., Spence P.R., Van Der Heide B. Social media as information source: Recency of updates and credibility of information J. Comp.-Med. Commun., 19 (2014), pp. 171-183.

[2] Rubin, V.L., Chen, Y., Conroy, N.J., 2015. Deception detection for news: Three types of fakes, in: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research

ISSN: 2233-7857 IJFGCN

in and for the Community, American Society for Information Science, Silver Springs, MD, USA. pp. 83:1–83:4.

[3] Chen, Y., Conroy, N.J., Rubin, V.L., 2015. News in an online world: The need for an ”automatic crap detector”, in: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, American Society for Information Science, Silver Springs, MD, USA. pp. 81:1–81:4.

[4] Masri, R., Aldwairi, M., 2017. Automated malicious advertisement detection using virustotal, urlvoid, and trendmicro, in: 2017 8th International Conference on Information and Communication Systems (ICICS), pp. 336–341.

[5] Qbeitah, M.A., Aldwairi, M., 2018. Dynamic malware analysis of phishing emails, in: 2018 9th International Conference on Information and Communication Systems (ICICS), pp. 18–24.

[6] Pogue, D., 2017. How to stamp out fake news. Scientific American 316, 24–24.

[7] Varalakshmi Konagala and ShahanaBano (2020). Fake News Detection Using Deep Learning: Supervised Fake News Detection Analysis in Social Media with Semantic Similarity Method. IGI Global, p1-12.

[8] Balmas, M., 2014. When fake news becomes real: Combined exposure to multiple news sources and political attitudes of inefficacy, alienation, and cynicism. Communication Research 41, 430–454.

[9] Aldwairi M., Hasan M., Balbahaith Z.Detection of drive-by download attacks using machine learning approach Int. J. Inf. Sec. Priv., 11 (2017), pp. 16-28.

[10] Aldwairi, M., Al-Salman, R., 2011. Malurls: Malicious urls classification system, in: Annual International Conference on Information Theory and Applications, GSTF Digital Library (GSTF-DL), Singapore.

[11] Al Messabi, K., Aldwairi, M., Al Yousif, A., Thoban, A., Belqasmi, F., 2018. Malware detection using dns records and domain name features”, in: International Conference on Future Networks and Distributed Systems (ICFNDS), ACM.

[12] Abu-Nimeh, S., Chen, T., Alzubi, O., 2011. Malicious and spam posts in online social networks. Computer 44, 23–28.

[13] Brewer P.R., Young D.G., Morreale M. The impact of real news about fake news”: Intertextual processes and political satire International Journal of Public Opinion Research, 25 (2013), pp. 323-343

[14] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, Michael M. Bronstein (2019). Fake News Detection on Social Media using Geometric Deep Learning, p1-15.

[15] EtharQawasmeh, MaisTawalbeh and Malak Abdullah (2019). Automatic Identification of Fake News Using Deep Learning. Sixth International Conference on Social Networks Analysis, Management and Security, IEEE, p383-388.

[16] Kaggle Fake News Dataset. Retrieved from https://www.kaggle.com/c/fake-news/data.

[17] Y.Md.Riyazuddin,G.Susmithavalli, G.Victor Daniel.An Approach For The Implementation Of Double Guard MethodFor Detecting The Ids In Web ApplicationsInternational Journal Of Advanced Research In Management, Architecture, Technology And Engineering 2018, Vol 4(6)

[18] Susmitha Valli Gogula, Karthik Rajendra, "A Novel Approach For Noise Filtering From Diabetic Retinopathy Images Using Improved Pillar K-Means Algorithm", Journal Of Fundamental And Applied Sciences Issue 10(6s).

[19] B. Madhuravani, D. S. R. Murthy, S. V. Raju, “AN IMPROVED WIRELESS NODENEIGHBOR INTEGRITY VERIFICATION AND ENCRYTION USING ADDITIVE AND

ISSN: 2233-7857 IJFGCN

MULTIPLICATIVE HOMOMORPHIC MODEL", Journal of Fundamental and Applied Sciences, ISSN 1112-9867, 2018, 10(6S), 2911-2932.

[20] B. Madhuravani, B. Rama, N. Chandra Sekhar Reddy, B.Dhanalaxmi, V.UmaMaheswari, "A MATHEMATICAL MODEL OF INTEGRATED CHAOTIC BASED HASH FUNCTION TO IMPROVE RANDOMIZATION AND BIT VARIATION PROPERTIES OF CHAOTIC MAPS", J.Mech.Cont.&Math.Sci., Vol.13, No.-5, November-December(2018) Pages 225-240

[21] K. Sai Prasad, Dr. S Pasupathy, " Deep Learning Concepts and Libraries Used in Image Analysis and Classification", TEST Engineering & Management, Volume 82, Page number 7907-7913, January-February 2020.

ISSN: 2233-7857 IJFGCN