FAKE NEWS DETECTION USING PYTHON AND MACHINE LEARNING

PRESENTER NAME:

- NAME 1: HARIHARAN S
- DEPARTMENT OF ORGANIZATION: COMPUTER SCIENCE ENGINEERING
- NAME OF ORGANIZATION:RAJALAKSHMI INSTITUTE OF TECHNOLOGY
- EMAIL:hariharan.s.2021.cse@ritchennai.edu.in
- NAME 2: ELANGOVAN K
- DEPARTMENT OF ORGANIZATION: COMPUTER SCIENCE ENGINEERING
- NAME OF ORGANIZATION:RAJALAKSHMI INSTITUTE OF TECHNOLOGY
- EMAIL:elangovan.k.2021.cse@ritchennai.edu.in
- NAME 3 : DHARSHAN Y
- DEPARTMENT OF ORGANIZATION: COMPUTER SCIENCE ENGINEERING
- NAME OF ORGANIZATION:RAJALAKSHMI INSTITUTE OF TECHNOLOGY
- EMAIL:dharshan.y.2021.cse@ritchennai.edu.in

INTRODUCTION

- Fake news, which is misleading material disseminated as news, has been on the rise since 2010, when Social Network Services (SNSs) like Facebook and Twitter first gained popularity. It became a major topic and had a big impact on people's voting choices in the 2016 US Presidential Election.
- [1] During the election, fake news on Facebook was primarily utilised to favour a particular candidate.
- [2]. Global mainstream media came together to offer users a confidence index for stories and hired individuals to keep an eye out for fake news to stop its spread
- [3]. Additionally, there have been numerous attempts to use a technical solution to resolve this issue.
- FOR EXAMPLE:
- There are, for instance, detection techniques based on artificial intelligence (AI) and techniques that identify the unusual dispersion pattern of the spread of false information [4]. Models that have been trained on data are used in AI-based detection techniques; this technique is categorised as a machine learning task for natural language processing (NLP). Several earlier studies, utilising decision trees or neural network models, achieved >80% accuracy [5, 6].

PROCESS TO FIND FAKE NEWS

- In fact, the use of machine learning for fake news identification is fascinating. Please be aware that while I can give you a broad overview of how you can approach this issue using Python and machine learning techniques, developing a thorough and precise fake news detection system can be a challenging and continuing research topic.
- 1.Dataset: Obtain a dataset of news items with labels designating which ones are "real" or "fake" news. Online resources include a number of datasets, including the LIAR dataset and the Fake News Challenge dataset.
- 2. Text Preprocessing: Give the news articles a text preprocessing. In this step, the text is cleaned up by deleting extra characters or HTML elements, as well as tokenization, stop word removal, stemming, and lemmatization.
- 3.Feature Extraction: Create numerical features from the preprocessed text that machine learning algorithms can process. Bag-of-words (BOW), TF-IDF, and word clouds are a few methods for feature extraction that are frequently employed.

- 4.Model Selection: Select a machine learning model that can accurately determine if news stories are authentic or not. For text classification problems, Naive Bayes and Support Vector Machines (SVM) are two popular models.
- 5.Model Training: Separate the training and test sets from your labelled dataset. To train the machine learning model of your choice, use the training data. In this step, the model's parameters are optimised to reduce prediction errors while also including the extracted features and associated labels.
- 6.Model Evaluation: Utilising the testing set, assess how well the trained model performed. A few common evaluation metrics are F1-score, recall, accuracy, and precision. You may also examine the confusion matrix to comprehend how well the model does at differentiating between bogus and legitimate news

- 7.Deployment: Once you are satisfied with the performance of your model, you can use it to categorise fresh, unused news pieces. Users can enter a news story into a straightforward user interface, and your model will determine if it is authentic or fraudulent.
- Remember that spotting fake news is difficult due to the continuously changing tactics used by those who spread misinformation. To increase the accuracy of your model, you might need to update and modify it frequently.Python offers a number of tools and frameworks, such as scikit-learn, NLTK, TensorFlow, or Keras, that can help you put these processes into practise. To aid in your development process, be sure to become familiar with these resources.

MODEL ARCHITECTURE

- The difficult job of detecting fake news can be handled using a variety of machine learning techniques. Using supervised learning algorithms and natural language processing (NLP) is one well-liked method. Here is a high-level description of a model architecture for Python-based machine learning-based fake news detection:
- 1.Data Preprocessing: The raw text data must first be processed. This usually entails eliminating punctuation, changing the text's case to lowercase, eliminating stop words (frequently used words like "and," "the," etc.), and using stemming or lemmatization to break down words into their simplest forms

- 2.Feature Extraction: The next step is to extract numerical features from the preprocessed text that machine learning algorithms can recognise. Typical NLP feature extraction methods include the following:Represent each document as a vector of word frequencies or presence/absence indicators using the bag-of-words (BoW) method. Term frequency-inverse document frequency (TF-IDF) matrix is a common name for the resulting matrix.Embeddings in Word To represent words as dense vectors in a continuous vector space, use pre-trained word embeddings like Word2Vec, GloVe, or FastText.N-grams: To capture more contextual information, include not only single words but also pairs or triplets of words (n-grams).Using methods like Latent Dirichlet Allocation (LDA) or Non-Negative Matrix Factorization (NMF), extract latent topics from the text.

- 3.Model Training: After the features have been retrieved, a machine learning model may be trained to determine if news articles are authentic or not. For this purpose, some frequently used algorithms are as follows:Naive Bayes is a probabilistic method that frequently performs text classification jobs and is based on the Bayes theorem.Support Vector Machines (SVM) is a supervised learning technique that divides data points into various groups using hyperplanes.An ensemble learning technique called Random Forest uses several decision trees to produce predictions.Gradient boosting refers to algorithms like XGBoost or LightGBM that create an ensemble of flawed weak models successively, with each model improving on the one before it.Deep Learning: The detection of fake news can also be done using neural network architectures like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs).

- 4.Model Evaluation: After the model has been trained, you must assess its effectiveness. Accuracy, precision, recall, F1 score, and area under the ROC curve (AUC-ROC) are common evaluation measures for classification tasks. To provide a more accurate estimate of the model's performance, cross-validation methods like k-fold cross-validation can be utilised
- 5.Model Deployment: The model can be used to make predictions on fresh, untested data after it has been trained and assessed. The model can be made available as an API or integrated into a web application to do this.
- It's important to note that the success of the fake news detection model relies on the quality and diversity of the training data, as well as the chosen features and machine learning algorithms. Additionally, regular updates and retraining of the model are crucial to adapt to evolving patterns of fake news.

EXPERIMENTS

- As a dataset, we employ 100K items that were crawled from the websites of times of india,ndtv news,hindustan times,business times,the hindu. We divide the news for each press into the categories of economics, society, politics, entertainment, and sports before gathering articles in the same ratio for each category. We use 31K for mission 1 and 68K for mission 2 of them. For each mission, the proportions of true and false news are the same, and the ratio of training to validation data is 9:1. We use test data from 350 current articles (as of March 2018) that are not part of the training and validation sets and are made up equally of true and false news to assess the model's accuracy.

EXPERIMENTAL RESULTS

- Table displays the results of our accuracy measurement using the model with the lowest validation loss among all of the model's steps.As a measurement technique, AUROC (area under receiver operating characteristic curve) is employed.
- **TABLE:**

model	mission 1	mission2
BCNNN	0.527	0.721
BiT-BCNNN	0.418	0.713
BiT-LSTMN-BCNNN	0.413	0.707
AOS-BCNNN	0.212	0.726

CONCLUSION

- The primary contributions of this paper, which constructs a deep learning model for fake news identification and assesses the accuracy, are as follows:
- (1) With an AUROC score of 0.726, APS-BCNN's classification accuracy for mission2, which consists of false information unrelated to the context of the article, is the highest. It can be inferred that the similarity vector between the title and body helps identify content that is out of place in the current context.

- (2) The accuracy of classification for mission1, which consists of fake news where the headline and (b) The quantity of training data varying between missions 1 and 2 would have resulted in a variation in accuracy between missions; nonetheless, we were able to acquirebody are inconsistent, is the highest with a BCNN in AUROC score of 0.52; however, thisaccuracy cannot be used to detect real fake news. We can deduce the causes of low accuracy asfollows: (a) as CNN uses the local information of texts to classify, mission2 would have achievedhigh accuracy due to the large amount of perturbed local information. However, as mission1 hasa relatively small amount of perturbed local information, it would have been difficult to classify body are inconsistent, is the highest with a BCNN in AUROC score of 0.52; however, thisaccuracy cannot be used to detect real fake news. We can deduce the causes of low accuracy it. (b) The difference in the amount of training data between mission1 and mission2 would havecaused a difference in the accuracy between missions; we were able to acquire a large amount offake news data for mission2 by mixing parts of the bodies of several articles. However, since wehad to make the fake news data for mission1 individually, it was difficult to acquire a largeamount of fake news data like for mission2

- (3) CNN with LSTM performs poorly for classification. Although LSTMCNN has previously been used to the text classification of a single input with good accuracy [19], the application of LSTM to the text classification of two inputs as demonstrated in this paper had low accuracy. We can infer the following factors as the cause of the poor accuracy: Assuming, for instance, that the word "apple" appears in both the headline and the body as given in Table
- **TABLE EXAMPLE:**
- EXAMPLE
- Headline This **apple** is too tasty
- Body The red **apple** on the desk seems so tasty
- In this paper, a useful deep learning model for fake news identification is proposed. The study's main shortcoming is that, while we were able to classify items with meaningful accuracy when the body's content was unrelated to the headline's context, accuracy suffered when the two were inconsistent

ACKNOWLEDGEMENT

- This paper is created by a intel team **TEAM TECHIE,** here we have explained the fake news detection by using python and machine learing
- All code and data is available at **GIT HUB**
- **LINK: https://github.com/Hariharansridharan/intel**

REFERENCE

- This paper is about the detection of fake news using python and machine learning
- As, our team have taken some reference to complete this paper presentation
- [1]. "Survey of fake news detection technology," IITP Weekly Trend, vol. 1816, pp. 12-23, 2017. Y. Yoon, T. Eom, J. Ahn, H. Lee, and J. Heo
- [2]. "News with integrity," The Trust Project [Online]. Easily accessible at https://thetrustproject.org.
- [3].Distributed representations of words and phrases and their compositionality: T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Advances in Neural Information Processing Systems, vol. 26, pp. 3111–3119, 2013.
- [4].“Evaluating Machine Learning Algorithms for Fake News Detection" by Shlok Gilda in the 2017 IEEE 15th Student Conference on Research and Development (SCOReD), December 2017, pp. 110–115
- [5].Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi, "Faking Sandy: Characterising and Identifying Fake Images on Twitter During Hurricane Sandy", The 22nd International Conference on the World Wide Web's proceedings