

# **Project Report**

## **Fake News Detection using Python and Machine Learning**

Team Name:PSA

Team Members:

- 1) Mohamed Aslam M
- 2) Prethisha V
- 3) Sahana K

### **I. Introduction:**

Fake news refers to intentionally false or misleading information presented as news or factual reporting. It can take various forms, including fabricated stories, manipulated images or videos, misleading headlines, and biased or distorted reporting. The primary purpose of fake news is often to deceive and mislead readers, either for political, ideological, financial, or sensationalistic reasons.

The spread of fake news has become a significant concern in today's digital age. Fake news can have serious consequences on public opinion, political stability, and social harmony. Therefore, the development of effective methods to detect and combat fake news is of paramount importance. In this report, we present a study on fake news detection using machine learning techniques, specifically focusing on logistic regression.

### **II.Related Work:**

In this section, we review existing literature and research on fake news detection. We explore various approaches, such as natural language processing, feature engineering, and machine learning algorithms, employed by researchers to tackle this problem.

[1] used naïve Bayes classifier to detect fake news by Naive Bayes. This method was performed as a software framework and experimented it with various records from the Facebook, etc., resulting in an accuracy of 74%. The paper neglected the punctuation errors, resulting in poor accuracy.

[2] estimated various ML algorithms and made the researches on the percentage of the prediction. The accuracy of various predictive patterns included bounded decision trees, gradient enhancement, and support vector machine were assorted. The patterns are estimated based on an unreliable probability threshold with 85-91% accuracy.

[3] utilized the Naive Bayes classifier, discuss how to implement fake news discovery to different social media sites. They used Facebook, Twitter and other social media applications as a data sources for news. Accuracy is very low because the information on this site is not 100% credible.

[4] aimed to utilize machine learning methods to detect fake news. Three common methods are utilized through their researches: Naïve Bayes, Neural Network and Support Vector Machine (SVM). Normalization technique is an essential stage in data cleansing prior machine learning is used to categorizing the data. The output proved that that Naïve Bayes has an accuracy of 96.08% for detecting fake messages. Two more advanced methods, the neural network and the machine vector (SVM) reached an accuracy of 99.90%.

[5]presented a counterfeit detection model using N-gram analysis by the lenses of various characteristic extraction techniques. In addition, we examined the extraction techniques of various features and six different methods of machine learning. The proposed model achieves the highest accuracy in use ,Contains a uni gram and a linear SVM workbook. The highest accuracy is 92%.

### **III.Contributions:**

This report presents a machine learning model developed by our team for predicting fake and true news. The model was trained using the **ISOT Fake news Dataset**, which includes two separate datasets: Fake.csv and True.csv. These datasets were combined, and fake and true news were labeled as 0 and 1, respectively. The news data was then **cleansed** by removing numbers, square brackets, HTML tags, and URLs. Afterward, the news text was **vectorized** using TF-IDF Vectorizer, which calculates word frequencies and their importance in the entire dataset. Finally, the **Logistic Regression** algorithm was implemented to train the model using the vectorized news dataset.

Logistic Regression,a supervised algorithm,mainly used for binary classification ,was opted for its simplicity and its high performance in classification tasks.The trained model was later tested which unseen data,where the model achieved an accuracy of **98.83%**,which is higher than the models proposed in [1],[2],[3] and [5].In addition,the model was well trained to handle punctuation error,which wasn't observed in [1] which led to its low accuracy. [5] makes use of six machine learning algorithms for fake news detection were the highest accuracy observed from the six algorithms is 92% which is comparatively lesser than our model's accuracy of 98.83%.

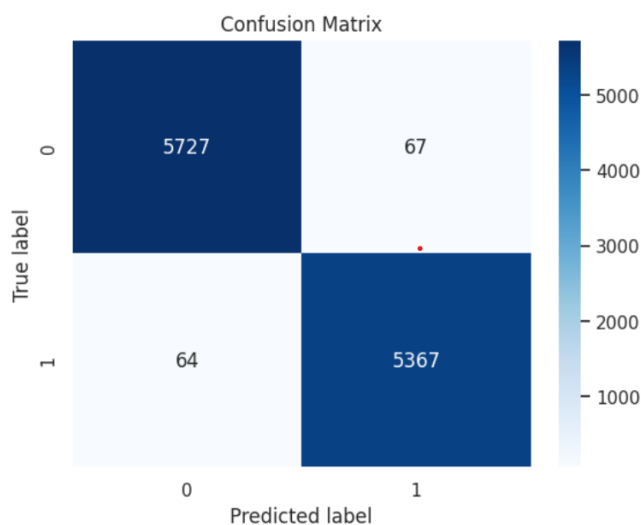
### **IV.Evaluation:**

The performance of the model was evaluated based on metrics such as confusion matrix,which shows the number of true and fake news prediction and classification report,which provides a detailed analysis on the metrics such as recall,F1-score,precision and accuracy\_score where the model achieved an accuracy of 98.83%

Classification Report:

Classification Report					
		precision	recall	f1-score	support
	0	0.99	0.99	0.99	5794
	1	0.99	0.99	0.99	5431
	accuracy			0.99	11225
	macro avg	0.99	0.99	0.99	11225
	weighted avg	0.99	0.99	0.99	11225

Confusion Matrix:



Accuracy:

```
[19]: #TODO: Model training and print the accuracy score
LR=LogisticRegression()
LR.fit(xv_train,y_train)
LR_prediction=LR.predict(xv_test)
accuracy=accuracy_score(y_test,LR_prediction)
print(accuracy)

0.9883296213808463
```

V. Future Works:

While our study yields promising results, there are several avenues for future research. We discuss potential areas for improvement, such as exploring alternative machine learning algorithms, incorporating deep learning techniques,ensemble learning techniques and leveraging additional data sources. We also suggest extending the model to handle multi-class classification, as fake news detection often involves classifying news into multiple categories of misinformation.

## **VI.Conclusion:**

In conclusion, our study demonstrates the effectiveness of logistic regression for fake news detection. The high accuracy achieved by our model underscores its potential as a valuable tool in combating the proliferation of misinformation. However, ongoing research and development are crucial to address the evolving nature of fake news and the ever-increasing sophistication of deception techniques. By continuing to refine and enhance our models, we can contribute to the broader effort of promoting information integrity and ensuring a more informed society.

## **VII.Reference Papers:**

- [1] Gilda, S. "Evaluating machine learning algorithms for fake news detection." *15th Student Conference on Research and Development (SCoReD)* (pp. 110-115). IEEE. 2017.
- [2] Akshay Jain and AmeyKasbe. "Fake News Detection." *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. Bhopal, India: IEEE. 2018.
- [3] Yumeng Qin et al. "Predicting Future Rumours." *Chinese Journal of Electronics* ( Volume: 27 , Issue: 3 , 5 2018, 514 - 520.
- [4] Prabhjot Kaur et al. " Hybrid Text Classification Method for Fake News Detection." *International Journal of Engineering and Advanced Technology (IJEAT)* , 2388-2392. 2019.
- [5]Looijenga, M. S. "The Detection of Fake Messages using Machine Learning." *29 Twente Student Conference on IT, Jun. 6th, 2018, Enschede, The Netherlands*. Netherlands: essay.utwente.nl. 2018.