



**Karunya INSTITUTE OF TECHNOLOGY AND SCIENCES**

(Declared as Deemed to be University under Sec.3 of the UGC Act, 1956)

MoE, UGC & AICTE Approved

**NAAC A++ Accredited**

*An internship report submitted by*

**STUDENT NAMES - ISHAN CHASKAR (URK21CS1181),  
JOEL RENNY(URK21CS7008),  
JESWIN KUNJUKUTTY JOHNSON  
(URK21CS1097)**

*in partial fulfillment for the award of the degree of*

**BACHELOR OF TECHNOLOGY  
in  
COMPUTER SCIENCE AND ENGINEERING**

*under the supervision of*

**MENTOR NAME:  
R.Venkatesan M.E., Ph.D.  
Associate Professor**



**DIVISION OF COMPUTER SCIENCE AND ENGINEERING  
KARUNYA INSTITUTE OF TECHNOLOGY AND SCIENCES**

(Declared as Deemed to be University under Sec-3 of the UGC Act, 1956)

**Karunya Nagar, Coimbatore - 641 114. INDIA**



**Karunya INSTITUTE OF TECHNOLOGY AND SCIENCES**

(Declared as Deemed to be University under Sec.3 of the UGC Act, 1956)

MoE, UGC & AICTE Approved

**NAAC A++ Accredited**

**DIVISION OF COMPUTER SCIENCE AND ENGINEERING**

**BONAFIDE CERTIFICATE**

This is to certify that the report entitled, “Fake News detection using Python and Machine Learning” is a bonafide record of Internship work done at INTEL during the academic year 2022-2023 by

**STUDENT NAME: ISHAN CHASKAR (URK21CS1181)**

**JOEL RENNY(URK21CS7008)**

**JESWIN KUNJUKUTTY JOHNSON**

**(URK21CS1097)**

in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering of Karunya Institute of Technology and Sciences.

---

**Guide Signature**

<Guide/Mentor Name>

<Designation>

## **ACKNOWLEDGEMENT**

First and foremost, I praise and thank ALMIGHTY GOD whose blessings have bestowed in me the will power and confidence to carry out my internship.

I am grateful to our beloved founders Late. Dr. D.G.S. Dhinakaran, C.A.I.I.B, Ph.D and Dr. Paul Dhinakaran, M.B.A, Ph.D, for their love and always remembering us in their prayers.

We extend Our tanks to Dr. Prince Arulraj, M.E., Ph.D., Ph.D., our honorable vice chancellor, Dr. E. J. James, Ph.D., and Dr. Ridling Margaret Waller, Ph.D., our honorable Pro-Vice Chancellor(s) and Dr. R. Elijah Blessing, Ph.D., our respected Registrar for giving me this opportunity to do the internship.

I would like to thank Dr. Ciza Thomas, M.E., Ph.D., Dean, School of Engineering and Technology for his direction and invaluable support to complete the same.

I would like to place my heart-felt thanks and gratitude to Dr. J. Immanuel John Raja, M.E., Ph.D., Head of the Division, Computer Science and Engineering for his encouragement and guidance.

I feel it a pleasure to be indebted to, R.Venkatesan M.E., Ph.D. Associate Professor, Division of CSE & Mrs. Nazneen Sultana AI Software Solutions Engineer AI & Analytics, Intel Corporation for their invaluable support, advice and encouragement.

I also thank all the staff members of the School of CST for extending their helping hands to make this in Internship a successful one.

I would also like to thank all my friends and my parents who have prayed and helped me during the Internship.

## **Report on Fake news detection using machine learning**

*Team Name: ALPHA*

*Team Members: ISHAN CHASKAR , JOEL RENNY , JESWIN JOHNSON*

*Name of the institution: Karunya Institute of Technology and Sciences*

### **1. INTRODUCTION:**

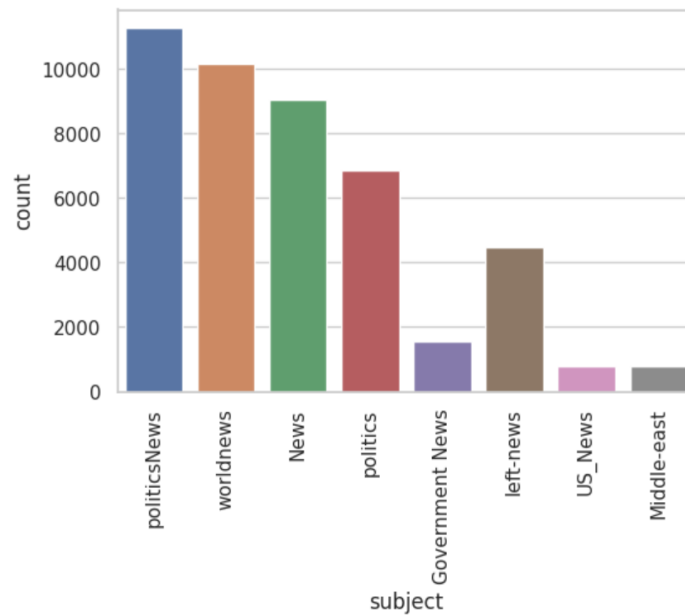
- **WHAT IS FAKE NEWS:** Fake news refers to false or misleading information presented as news or factual reporting. It involves the dissemination of fabricated or distorted content with the intention of deceiving or misleading readers or viewers. Fake news can take various forms, including articles, images, videos, or social media posts.

The primary characteristic of fake news is its lack of accuracy and verifiability. It often contains sensationalized or exaggerated claims, misleading headlines, manipulated images, or quotes taken out of context. Fake news can be created and spread for various reasons, such as political propaganda, financial gain, or to generate controversy or confusion.

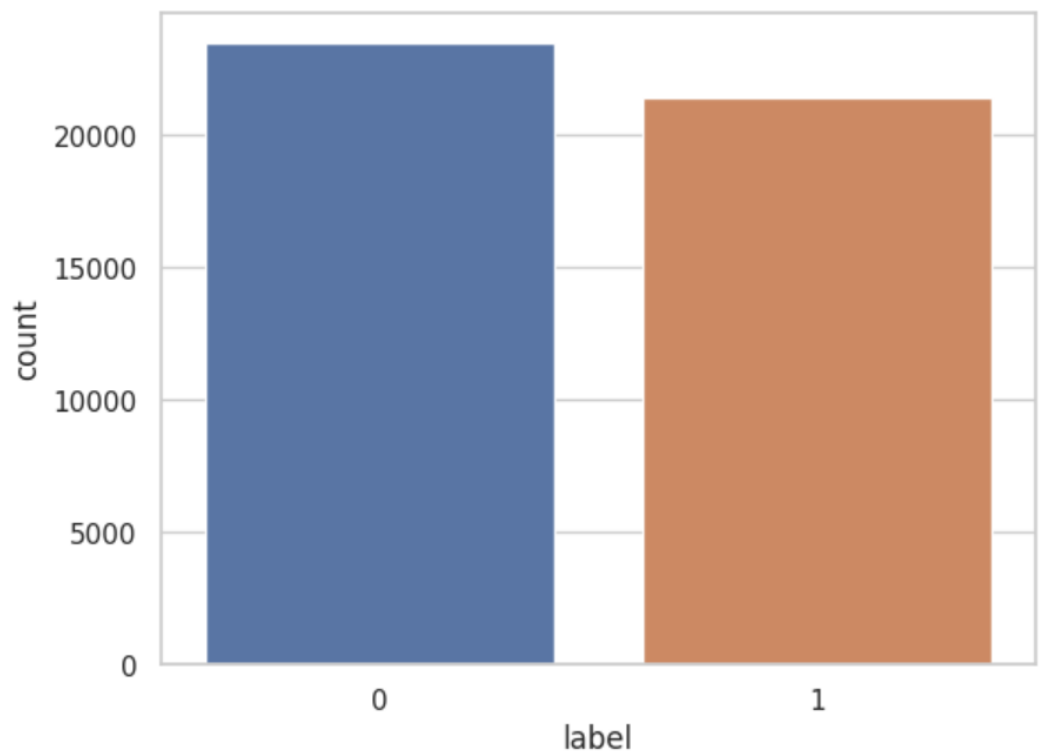
- **PROBLEM STATEMENT:** The problem statement of fake news detection revolves around developing effective methods and algorithms to identify and distinguish between genuine, reliable news and fabricated or misleading information. The goal is to create automated systems that can accurately detect and classify fake news, thereby helping individuals, organizations, and platforms make informed decisions and mitigate the negative impact of misinformation.
- **OBJECTIVE:** The primary objective of fake news detection is to develop accurate and reliable methods, models, and algorithms that can effectively identify and classify fake news from genuine, trustworthy information.
- **INTEL LIBRARIES:** Intel does offer optimized versions of popular Python libraries, including NumPy, SciPy, and sci-kit-learn, through the Intel Distribution for Python, there is no specific Intel-branded version of sci-kit-learn called "Intel scikit." The Intel Distribution for Python focuses on providing performance improvements by integrating with Intel Math Kernel Library (Intel MKL) and other optimized components.

### **2. EXPLORATORY DATA ANALYSIS (EDA):**

- **Differentiation based on types of news:** The count plot function is not specific to machine learning (ML) but rather a visualization function provided by various data analysis and visualization libraries, such as Seaborn in Python. It is commonly used to display the count or frequency of categorical variables in a dataset. While it can be useful in the context of data exploration and analysis, it does not have a direct association with ML algorithms. In this countplot is used to display the classification of the news based on different metrics

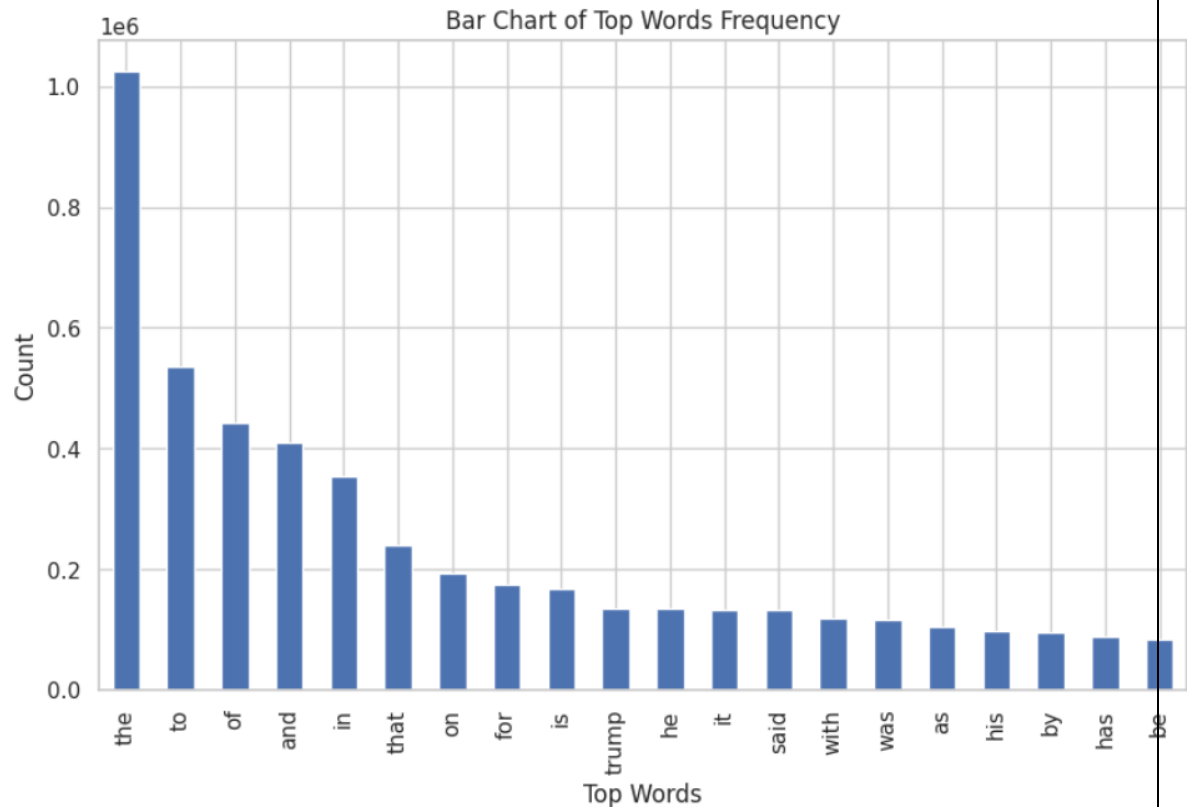


- *Count of fake and real news:* To count the occurrences of real and fake news within a dataset, you need to have a labeled dataset where each news article is categorized as either real or fake. Assuming you have such a dataset, you can use various programming techniques to count the occurrences.



- *Count of the top words:* To obtain the count of the top words in fake news detection, the text data is preprocessed by removing punctuation and stopwords. The preprocessed text is then tokenized into individual words. Word frequencies are calculated by counting the occurrences of each word. The words are sorted in descending order based on their frequency,

and the desired number of top words is selected. The count of the top words is obtained by retrieving their frequency counts.



- *Real-time prediction:* Real-time prediction in fake news detection involves deploying a trained machine learning model to quickly assess the authenticity of news articles as they are published or shared. The model is trained using historical data that includes labeled instances of real and fake news articles. Once deployed, the model receives new articles as input, processes the text data using preprocessing techniques, and applies the trained model to predict whether the article is real or fake. Real-time prediction enables prompt identification of potentially misleading or deceptive news articles, aiding in the timely dissemination of accurate information.

Enter the News for Prediction Transgender people will be allowed for the first time to enlist in the U.S. military starting on Monday as ordered by federal courts, the Pentagon said on Friday, after President Donald Trump's administration decided not to appeal rulings that blocked his transgender ban

Ensemble Prediction: Genuine News

## **MODELS IMPLEMENTED:**

### **Logistic Regression**

Logistic regression is a statistical modeling technique used to predict the probability of a binary outcome based on input features. It is particularly suitable for problems where the dependent variable is categorical, such as fake news detection (true or false). The logistic regression model applies the logistic function (sigmoid function) to estimate the probability of the binary outcome.

### **RandomForestClassifier**

RandomForestClassifier is a machine learning algorithm that belongs to the family of ensemble methods. It is a popular classification algorithm that combines multiple decision trees to create a powerful model. RandomForestClassifier is widely used for various tasks such as classification, regression, and feature selection.

### **GradientBoostingClassifier**

GradientBoostingClassifier is a popular machine learning algorithm that belongs to the family of ensemble methods, specifically gradient boosting. It is primarily used for classification tasks, where it combines multiple weak learners (decision trees) to create a strong predictive model. GradientBoostingClassifier is known for its high predictive accuracy and robustness against overfitting.

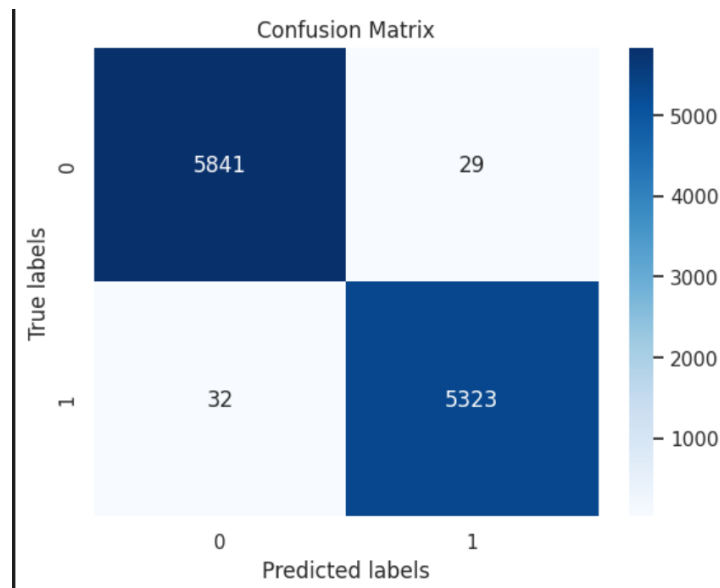
### **Voting Classifier**

Voting Classifier is a machine learning algorithm that combines the predictions of multiple individual classifiers to make a final prediction. It is a type of ensemble learning method that leverages the wisdom of the crowd to improve prediction accuracy and generalization.

## **RESULT OBTAINED:**

### **Confusion Matrix**

The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix itself can be easily understood, but the related terminologies may be confusing. Since it shows the errors in the model performance in the form of a matrix, hence also known as an error matrix.



### Ensemble Models

Ensemble models are machine learning techniques that combine the predictions of multiple individual models to make more accurate and robust predictions. Ensemble models leverage the diversity and collective intelligence of multiple models to overcome the limitations of individual models and improve overall performance.

Ensemble Accuracy: 0.9941202672605791

### Classification Report

It is one of the performance evaluation metrics of a classification-based machine learning model. It displays your model's precision, recall, F1 score and support. It provides a better understanding of the overall performance of our trained model.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	5870
1	0.99	0.99	0.99	5355
accuracy			0.99	11225
macro avg	0.99	0.99	0.99	11225
weighted avg	0.99	0.99	0.99	11225

### Model Evaluation

Model evaluation involves assessing the model's performance in distinguishing real and fake news articles. This includes selecting evaluation metrics, applying the model to a separate test set, calculating metrics like accuracy and precision, and interpreting the results. Comparative analysis, robustness testing, and visualization



techniques aid in understanding the model's behavior. Iterative refinement is performed based on evaluation results, ensuring continuous monitoring and improvement

#### Future Works

**Deep Learning Techniques:** Explore the use of advanced deep learning models like CNNs, RNNs, and transformers for more accurate and robust fake news detection.

**Multimodal Approaches:** Integrate multiple modalities such as text, images, videos, and social network data to improve the comprehensiveness of fake news detection.

**Explainable AI:** Develop techniques that provide interpretable explanations for the predictions made by fake news detection models to enhance transparency and user trust.