

# FAKE NEWS DETECTION



# AGENDA

---

1. Introduction
2. Dataset
3. Pre-processing and feature extraction
4. Methodology
5. Results
6. future work
7. Conclusion



# INTRODUCTION

---



# DATASET

---

Obtained From  
**kaggle**

Number of  
Samples  
**72,134**

Classified as  
**0 Real**  
**1 Fake**

Number of  
Features  
**3**

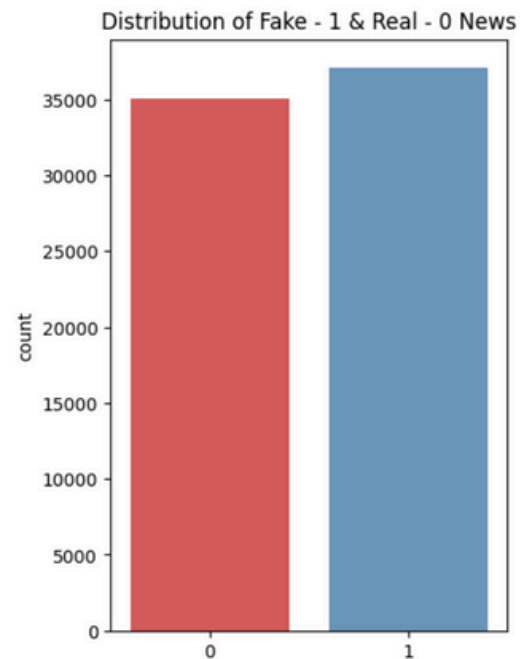
**Dataset Name : WELFake**



# DATASET

---

**Balanced dataset :**  
35,028 instances labeled as real  
37,106 instances labeled as fake



**FAKE  
NEWS**

## **PRE-PROCESSING AND FEATURE EXTRACTION**



**filling the null values**



**Stemming**

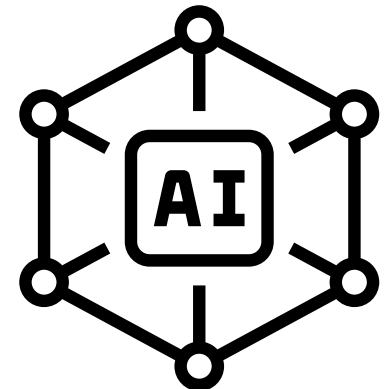


**TF-IDF (Term Frequency - Inverse  
Document Frequency)**

# METHODOLOGY

## Machine Learning Algorithms:

- Random Forest (RF)
- Support Vector Machine (SVM)
- Adaptive Boosting (AdaBoost)
- Extreme Gradient Boosting (XGBoost)
- Bagging Classifier
- Light Gradient Boosting Machine (LGBM)
- Gradient Boosting



# RESULTS

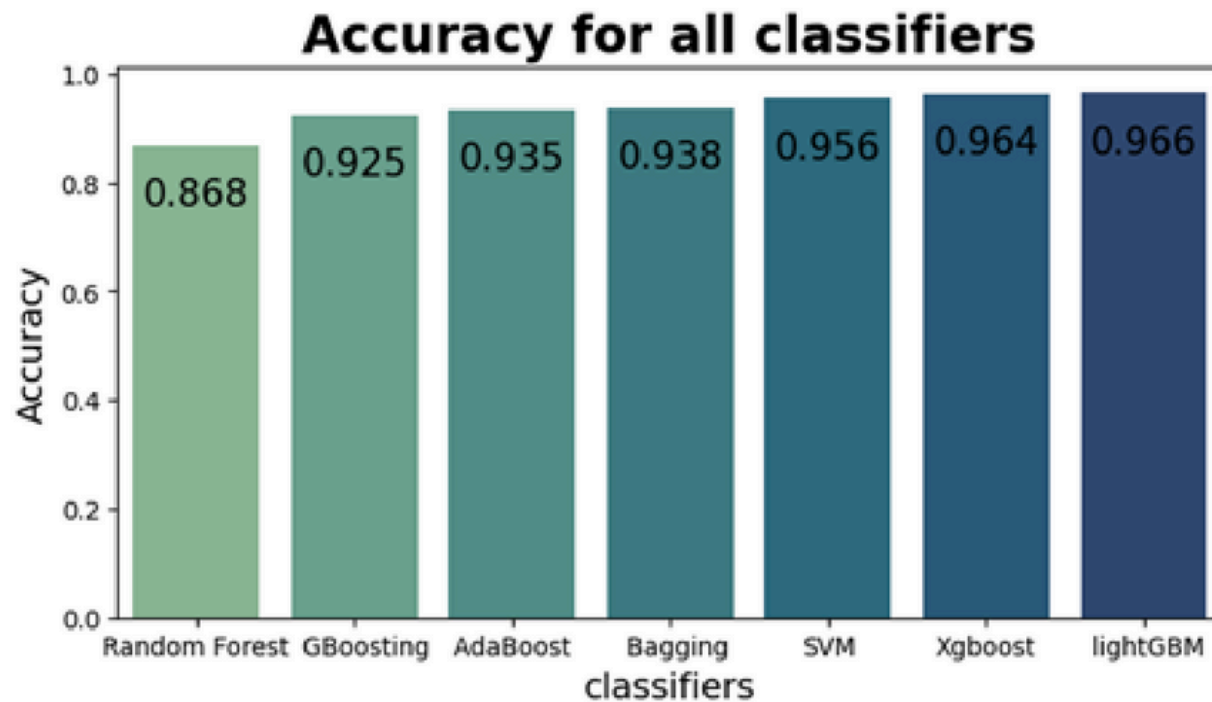
---

Classifier	Accuracy	Precision	Recall	F1-score
RF	86.8%	87%	87%	87%
SVM	95.6%	96%	96%	96%
AdaBoost	93.5%	94%	93%	94%
XGBoost	96.4%	96%	96%	96%
Bagging	93.8%	94%	94%	94%
lightGBM	96.6%	97%	97%	97%
GBoosting	92.5%	93%	92%	92%



# RESULTS

---



**Highest Accuracy : LGBM 96.6%**

# FUTURE WORK

---



Different machine learning algorithms



Real time implementation



# CONCLUSION

---





# THANK YOU FOR LISTENING



## Done by

- Abrar Abdulaziz
- Manar Alsayed
- Joury Alzayat
- Mona Alqahtani
- Noor Aljishi
- Ghaida Farhan