

SUMMER INTERNSHIP PROJECT

ON

FAKE NEWS DETECTION

SUBMITTED BY

SARVAGYA BANSAL

DATA SCIENCE JUNE BATCH



# ACKNOWLEDGEMENT

First and foremost, we would like to express my sincere gratitude to my teacher Mr.Mohit Tripathi, who gave me the golden opportunity to do this wonderful project on "FAKE NEWS DETECTION" and for their patience, motivation, enthusiasm, and immense knowledge. His guidance helped us in all the time of research and writing of this project.

Finally, We would like to thank my fellow classmates for many helpful discussions and good ideas along the way.

THANK YOU,  
SARVAGYA BANSAL



# WHAT IS FAKE NEWS ?

A type of yellow journalism, fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media.

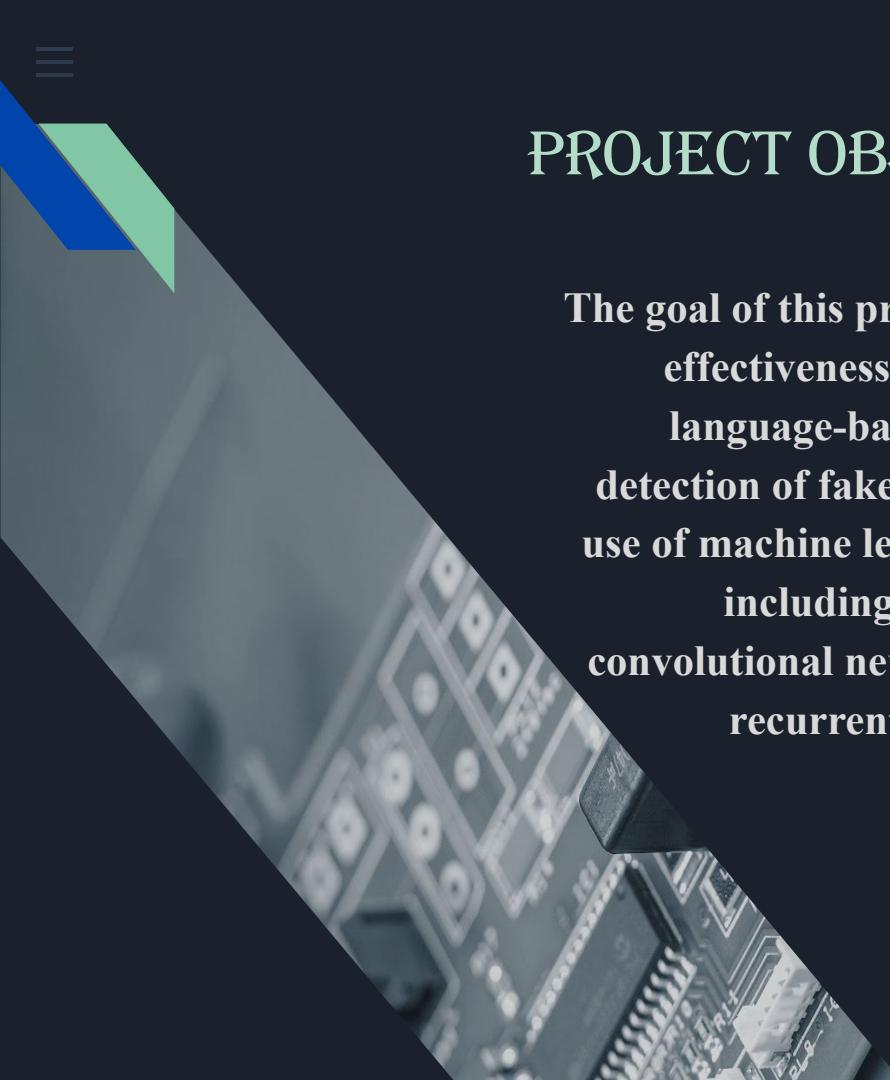
This is often done to further or impose certain ideas and is often achieved with political agendas.

Such news items may contain false and/or exaggerated claims, and may end up being viralized by algorithms.



# UNDERSTANDING THE PROBLEMS

- 01 First, Fake News Makes It Harder For People To See the Truth.
- 02 Second, they distract people from important issues so that these issues remain unresolved. This section explores how fake news is used for distraction and intensifying conflict.
- 03 Third, they intensify social conflict to undermine people's faith in the democratic process and people's ability to work together.



# PROJECT OBJECTIVE

**The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of machine learning algorithms including but not limited to convolutional neural networks and recurrent neural networks.**

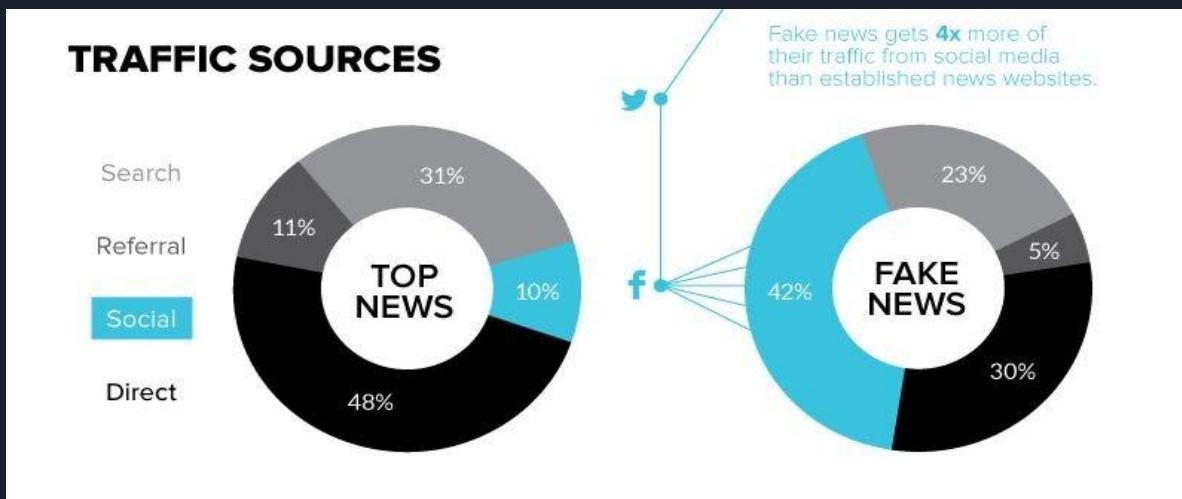
# TECHNOLOGY FOR FAKE NEWS DISSEMINATION



# SOCIAL MEDIA IS MAIN DRIVER FOR FAKE NEWS

10% of readers of top news come via social media

40% of readers of fake news come via social media





# WHAT IS A PASSIVE AGGRESSIVE CLASSIFIER?

Passive Aggressive algorithms are online learning algorithms. Such an algorithm remains passive for a correct classification outcome, and turns aggressive in the event of a miscalculation, updating and adjusting. Unlike most other algorithms, it does not converge. Its purpose is to make updates that correct the loss, causing very little change in the norm of the weight vector.

# ABOUT DETECTING FAKE NEWS WITH PYTHON -

This advanced python project of detecting fake news deals with fake and real news. Using **sklearn**, we build a **TfidfVectorizer** on our dataset. Then, we initialize a **Passive-Aggressive Classifier** and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.





# WHAT IS A TFIDFVECTORIZER?

**TF (Term Frequency):**

**The number of times a word appears in a document is its Term Frequency. A higher value means a term appears more often than others, and so, the document is a good match when the term is part of the search terms.**

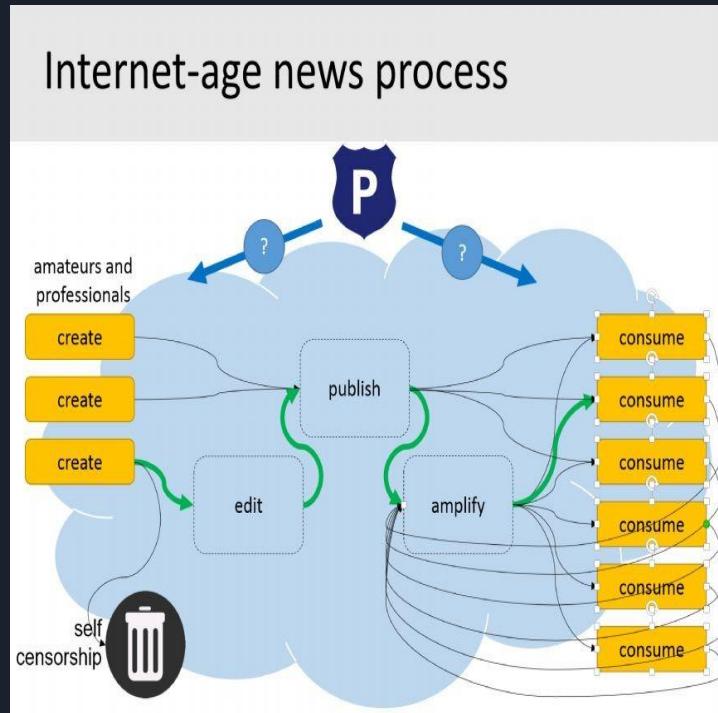
**IDF (Inverse Document Frequency):**

**Words that occur many times a document, but also occur many times in many others, may be irrelevant. IDF is a measure of how significant a term is in the entire corpus.**

# CAN WE STOP FAKE NEWS?

Yes we can!

- With Author
- With Publisher
- With Platform
- With Amplifier
- With Reader



Thank You!



Sarvagya Bansal  
Data Science June Batch