

# Exploring An Efficient Approach To News Classification

By

Team Machel

## Introduction

News is reporting about current events. We may think of it as a daily journal. Curated news about a certain topic over a period of time may become the only source of referential history on that topic to future generations.

In today's digital age, people are bombarded with news from a variety of sources. This can make it difficult to find the news that is relevant to our interests and to identify misinformation and fake news.

What should researchers type in their search for information about the tours Queen Elizabeth made in Nigeria?

For easy accessibility and referencing, there is a need to classify news into their appropriate sections. For example, different platforms report news using their own context; the way the government will report a communal clash in a region may differ from the local version. Similarly, events may involve multiple domains; an athlete involved in a car crash may be reported under the sporting category.

## Problem Description

Our goal is to identify the type of news based on headlines and short descriptions.

Why is it important to solve this problem?

1. Identify mis-information and fake news.

Fake news can be used to spread misinformation about elections, public health, and other important topics. It is a serious problem that can mislead people and have a negative impact on society. Identifying the type of news can help to identify fake news articles and prevent people from being misled.

A social media platform could use it to recommend news articles to users based on their interests. This would help users to discover new news sources and learn about new topics.

## 2. Support journalists and media agencies

The randomness with events occurs and the volume of content produced per day overwhelms media agencies. There is a need to organize their content, monitor news events, and generate reports efficiently.

Automated news classification can help save time and resources. These systems enhance efficiency, accuracy, and the ability to deliver high-quality content to an audience that relies on trusted news sources for accurate and timely information.

## 3. It can help to improve the accuracy of news recommendation systems.

Many news websites and platforms use recommendation systems to suggest articles to users. Effective news classification is crucial for providing users with content that matches their preferences and expectations.

## Proposed Method

1. **Begin with Training:** Initiate the project by training a model from scratch using classical machine learning methods and open source data. This foundational step will establish a benchmark for performance and help you gauge the effectiveness of these methods for the given task.

2. Local Data Acquisition: Collect data from Nigerian news websites through web scraping and assess the impact of training and testing on different data distributions on model performance. This analysis will shed light on the advantages of incorporating local data and how it influences the model's performance under varying data scenarios.
3. Architectural Exploration: Dive into more advanced model architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers. These sophisticated architectures have demonstrated remarkable performance in various text classification tasks and warrant exploration for your project.
4. Leverage Pre-trained Models: Evaluate the performance of pre-trained models from Hugging Face, such as BERT. These models are pre-trained on extensive text and code datasets and can be fine-tuned for diverse tasks, including news classification. Assess their effectiveness in enhancing your model's performance.
5. Deployment and Accessibility: Take the project to the next level by deploying the model as an API and crafting a user-friendly browser plugin. This plugin will allow real-time content scraping from news websites and provide instant inference, making the model accessible and valuable to a wider audience.

## Proposed Split

Data Sourcing Sub team: Mustapha Rufai, Akojede Olorundara, and Bilikis

Data Cleaning and Preparation sub team: Bilikis and Si

Model Sub team: Mustapha Rufai, Akojede Olorundara

Model Deployment sub team: Mustapha Rufai, Akojede Olorundara, and Bilikis

## Conclusion

News classification is the task of assigning news articles to predefined categories, such as business, sports, entertainment, or politics. This task can be challenging due to the large volume of news articles published daily, the variety of topics covered, and the complexity of natural language.

## References

1. Misra, Rishabh. "News Category Dataset." arXiv preprint arXiv:2209.11429 (2022).
2. Misra, Rishabh and Jigyasa Grover. "Sculpting Data for ML: The first act of Machine Learning." ISBN 9798585463570 (2021).
3. Saravanou, A.; Panagiotou, N.; Gunopulos, D. News Monitor: A Framework for Querying News in Real Time. In Proceedings of 43rd European Conference on IR Research, ECIR 2021, Virtual Event, 28 March–1 April 2021; pp. 543–548”.

[https://www.researchgate.net/publication/350476015\\_News\\_Monitor\\_A\\_Framework\\_for\\_Querying\\_News\\_in\\_Real\\_Time](https://www.researchgate.net/publication/350476015_News_Monitor_A_Framework_for_Querying_News_in_Real_Time)

### Mentor

- Kenechi Dukor

### Members

- [Akojede Olorundara](#)

- Bilikis
- Mustapha Rufai
- Si