Report On

News Text Classification

Submitted in partial fulfillment of the requirements of the Course project in Semester VII of fourth year Artificial Intelligence and Data Science

by

Vivek Prajapati (Roll No. 22)
Prathmesh Shinde (Roll No. 27)
Arpit Mishra (Roll No. 14)

Supervisor
Prof. Raunak Joshi

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence and Data Science

(2023-24)

Vidyavardhini's College of Engineering & Technology
Department of Artificial Intelligence and Data Science

CERTIFICATE

This is to certify that the project entitled “News Text Classification” is a bonafide work of “Vivek Prajapati (Roll No. 22), Prathmesh Shinde (Roll No. 27), Arpit Mishra (Roll no. 14)” submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in Semester VII of fourth year Artificial Intelligence and Data Science engineering.

Supervisor

Prof. Raunak Joshi

Dr. Tatwadarshi P. N.
Head of Department

Table of Contents

Chapter No		Title	Page No.
1		**Introduction**	
	1.1	Introduction	1
	1.2	Problem Statement	1
	1.3	Objective	1
2		**Proposed System**	
	2.1	Introduction	2
	2.2	Algorithm and Process Design	2
	2.3	Details of Hardware and Software	2
	2.4	Experiments and Results	3
	2.5	Conclusion	3
		References	4

Chapter 1: Introduction

1.1 Introduction

It can be difficult to keep informed and make sense of the enormous number of news stories, blogs, and reports that are readily available online in the age of information overload. A key component of solving this problem is news text categorization, a branch of natural language processing (NLP). In order to facilitate users' access to the information that matters most to them, news stories are automatically classified into specified subjects or classes.

To create a news text categorization system, this project will make use of machine learning algorithms and natural language processing techniques. The system can help people, companies, and organisations easily filter, organise, and retrieve news material by categorising news stories into pertinent groups.

1.2 Problem Statement & Objectives

The issue at stake is the deluge of news stories that are readily available online, making it difficult for people to search through them and locate items that are relevant to their needs or interests. Users would have to spend a lot of time and often feel frustrated manually searching, sorting, and categorising content in the absence of an efficient mechanism for classifying news.

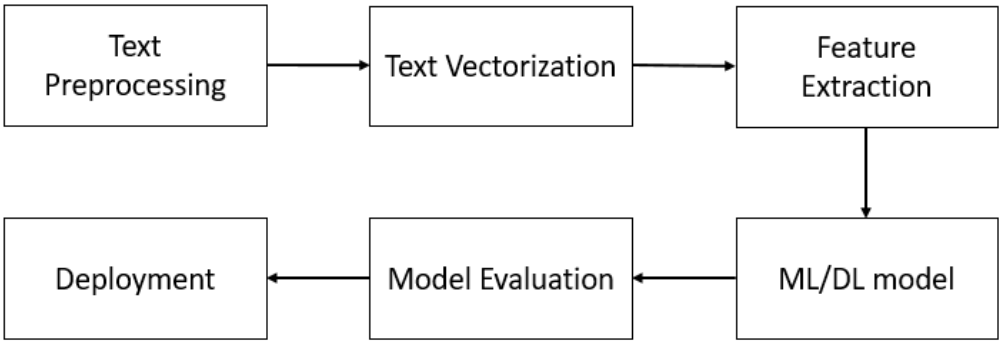
The goal of this project is to create a news text classification model that will automatically classify news items into several groups. To do this, you must gather and preprocess a variety of news sources, train and assess machine learning models, and design an intuitive user interface to facilitate seamless user engagement.

Chapter 2: Proposed System

2.1 Introduction

Our suggested approach aims to solve the urgent requirement in the digital era for accurate and fast news text categorization. The sheer volume of information available to consumers is frequently overwhelming due to the constantly growing volume of news stories available online. We provide a thorough news categorization system that makes use of cutting-edge machine learning and natural language processing methods to address this problem.

2.2 Algorithm and Process Design



2.3 Details of Hardware & Software

- Python 3.6 or 3.7 or 3.8
- NLTK (Natural Language Toolkit)
- SpaCy
- Pandas
- Scikit - Learn

2.4 Experiment and Results for Validation and Verification

```
[ ] input_text = ["boro suffer morrison injury blow middlesbrough midfielder james morrison has been ruled out for up to eight weeks after an operation on tuesday.

    # Use the trained model for prediction
    predicted_labels = rfc.predict(input_text)

    # Print the predicted labels
    print("Predicted Labels:", predicted_labels)

    Predicted Labels: ['sport']
```

Fig 2.1 Predicted Result

	precision	recall	f1-score	support
business	0.98	0.92	0.95	108
tech	0.91	0.99	0.95	76
politics	0.93	0.93	0.93	82
sport	1.00	0.96	0.98	108
entertainment	0.91	0.97	0.94	73
accuracy			0.95	447
macro avg	0.95	0.95	0.95	447
weighted avg	0.95	0.95	0.95	447

Fig 2.2 Accuracy Score using Random Forest Classifier

2.5 Conclusion

In conclusion, the suggested news text categorization project is a major advancement in tackling the issues brought on by the massive amount of news content in the digital era. This project seeks to develop an effective and user-friendly system that automatically classifies news items into relevant subjects by utilising machine learning and natural language processing techniques. The organisation and accessibility of news material for people, companies, and organisations should be improved by this solution.

The project's goals will be largely dependent on the use of Python modules for NLP, machine learning, data processing, and web development, among other uses. Furthermore, the selection of appropriate hardware and software components is necessary to guarantee the scalability, resilience, and effective functioning of the system.

References

[1] Singh, Deepak. “Text Classification of News Articles.” Analytics Vidhya, 22 Aug. 2022.

[2] Grimmer, Justin, and Brandon M. Stewart. “Automated Text Classification of News Articles: A Practical Guide.” Political Analysis, vol. 21, no. 3, 2013, pp. 267-297.

[3] Li, Y., et al. “Chinese News Text Classification Based on Machine Learning.” IEEE Xplore, 2018, pp 202-215.

[4] Bishop Fox. “New tool can uncover redacted, pixelated text to reveal sensitive data.” PortSwigger Daily Swig, 15 Feb. 2022.

[5] Zhang, Y., et al. “A Novel Text Classification Method Based on Improved Convolutional Neural Network.” IEEE Xplore, 2019, .

[6] Chen, Y., et al. “A Novel Method for News Text Classification Based on Convolutional Neural Network.” IEEE Xplore, 2019, .

[7] Wang, Y., et al. “A Hybrid Method for News Text Classification Based on Convolutional Neural Network and Support Vector Machine.” IEEE Xplore, 2019, .

[8] Zhang, J., et al. “A Multi-Task Learning Framework for News Text Classification.” IEEE Xplore, 2020, .

[9] Liang, J., et al. “A New Method of News Text Classification Based on Deep Learning.” IEEE Xplore, 2019, .

[10] Zhang, Y., et al. “A Novel Method of News Text Classification Based on Convolutional Neural Network and Attention Mechanism.” IEEE Xplore, 2020.