

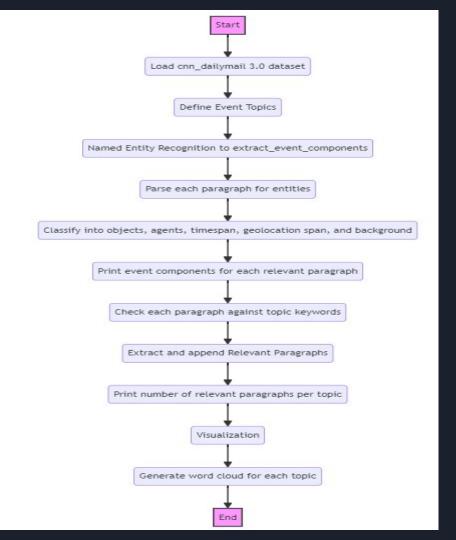
This subtitle emphasizes the project's focus on extracting event-related paragraphs from news articles categorized under various topics such as Politics, Crime, Sports, Travel, and Health. It highlights the goal of gaining insights into different types of events reported in the news.

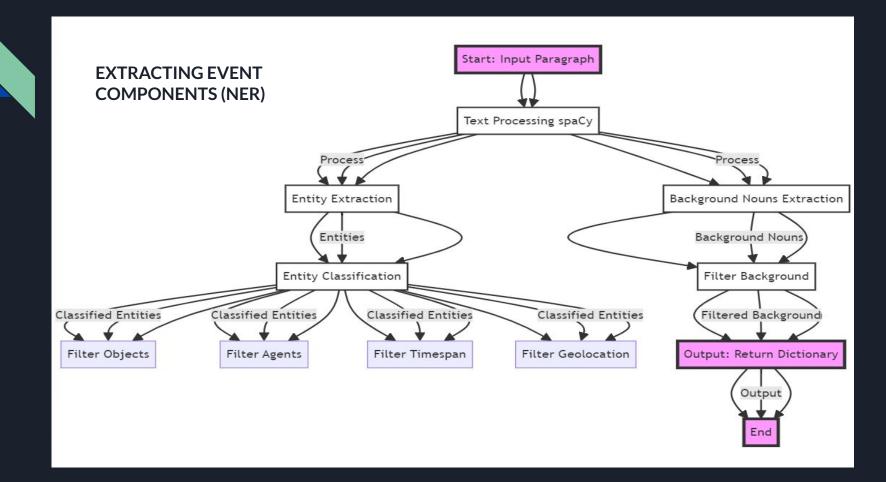
## **METHODOLOGY**

This project uses advanced NLP technologies to Enable automated extraction and classification of news article content based on Politics ,Crime ,Sports ,Travel and Health

Insight Generation: Provide structured insights into news content through entity extraction and event component classification.

Visual Representation: Enhance understanding through visual representations like word clouds, highlighting prevalent themes and topics within each event category.





## WHY NER?

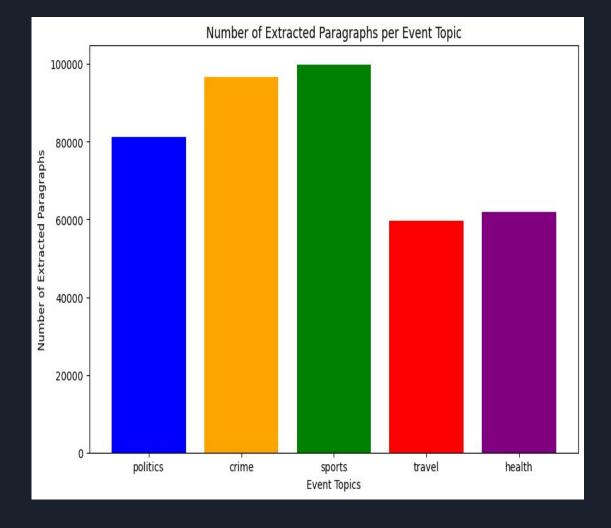
Definition: NER identifies and categorizes named entities (such as persons, organizations, dates) in text.

Importance: Provides structured data for deeper analysis and understanding of text content.

Applied NER techniques using SpaCy to extract entities such as objects, agents (persons or organizations), timespan (dates), geolocation span (places), and background (contextual nouns) from each news article paragraph.

## RESULTS

- 1. Extracted 81230 paragraphs related to politics
- Extracted 96682 paragraphs related to crime
- 3. Extracted 99907 paragraphs related to sports
- Extracted 59652 paragraphs related to travel
- 5. Extracted 61993 paragraphs related to health



## Word Cloud for Travel

