**Project Objective: Topic Modeling and Content Classification for News Articles**

**Data Sources:**

1. News articles from various online news portals.
2. Social media posts and discussions related to news topics.

**Exploratory Data Analysis (EDA) Questions:**

1. What are the dominant topics present in the news articles dataset?
2. How does the distribution of article lengths vary across different topics?
3. Are there any temporal patterns in the emergence of specific news topics?
4. What are the most frequently mentioned entities (e.g., people, organizations) in the articles?
5. How do sentiment scores vary within different news categories?
6. Are there common phrases or terms associated with specific news topics?
7. What is the geographic distribution of news coverage for different topics?
8. How diverse is the vocabulary across different news categories?
9. Are there correlations between the popularity of news topics and social media discussions?
10. How reliable and consistent are the assigned news categories or labels?

**Modelling Questions:**

1. Which NLP techniques are suitable for topic modeling in the context of news articles?
2. How can the model distinguish between different types of news (e.g., political, entertainment)?
3. What approaches can be used to handle multi-label classification for news articles covering multiple topics?
4. How will the model handle bias or subjectivity in news articles?
5. What preprocessing steps are necessary for effective topic modeling and content classification?
6. How can the model adapt to evolving news language and emerging topics?
7. What evaluation metrics are most relevant for assessing the performance of the topic modeling system?
8. How will the model deal with articles that are ambiguous or cover multiple distinct topics?
9. Can the model identify and filter out fake or misleading news articles?
10. How will user feedback and corrections be incorporated into the topic modeling system?

**Valuation Questions:**

1. How can topic modeling and content classification improve user engagement with news content?
2. What is the potential impact of accurate content categorization on user satisfaction?
3. How will the topic modeling results be integrated into news recommendation algorithms?
4. What is the cost-benefit analysis of implementing and maintaining the topic modeling system?
5. How can the system contribute to better understanding and addressing information biases?
6. What partnerships or collaborations could benefit from accurate news topic identification?
7. How will the model adapt to changes in news consumption patterns and public interests?
8. What additional features or insights can be derived by combining topic modeling with sentiment analysis?
9. How will the accuracy of content classification be measured and communicated to stakeholders?
10. How can the topic modeling system enhance the overall quality and credibility of news coverage?