Problem Statement:

Kavida.ai is a supply-chain disruption decision platform. We build a digital twin (replica) of enterprise supply chains and use artificial intelligence to detect disruption threats before they occur. Our AI then simulates the threat's impact and the optimal resiliency decision to minimize the financial and consumer impact with a mission to protect supply chains with integrated data so companies can proactively manage disruption risks. With the evolving business scope, it is crucial to set up processes that can help in delivering intelligent data-driven solutions.

Given you have joined the team as a Data Science Intern to help supply quality "Insights!"

You have been introduced to a vital challenge, a few of our clients want data for specific niche like alerts which might be related to environment/strike etc. related issues. As we proceed to scale the data offerings, we need an intelligent model to classify the articles which outperforms regular keywords-based classification!

TASK 1:

Pls find the attached live data from our systems which contain thousands of news articles and their headline with the corresponding categories.

Use transfer learning to fine-tune a classification model for the attached dataset.

Follow the guidelines below for this task,

- Focus on the proper *EDA of the dataset, and comment on your views!*
- *For model training you can use Kaggle or collab or personal GPU.*
- *Accuracy is not the concern, hence please limit your training time, the process followed by the rationale is more critical.*
- *apart from the code and proper comment, provide the rationale behind the models you use.*
- *Evaluate your trained model and provide the rationale behind the metrics used.*

TASK 2:

The intelligent model we created for the above problem is not valuable unless we deploy it to provide insights from the live data!

Create a RESTful API to get the inference from the model!

- Docker deployment (Recommended):
  - Containerize the application to get the inference from the model.
  - The request should respond with the predictions in a Json format.

(It is "optional" to make an end-to-end deployment on any of the cloud service)

TASK 3:

As a Data Scientist you will deal with a lot of problems related to data sourcing and distribution, what will be your approach to deal with the following:

1. For building a classification model for example for categorizing data into various genres like sports, science, etc., how would you get labeled data for this task? How would you collect such data from the internet?

2. How can you leverage the attached data and trained classification model (Task 1) to inform companies about supply chain disruptions? Please illustrate using a visualization.