# PROJECT REPORT: ZERO-SHOT TOPIC CLASSIFICATION

## 1. INTRODUCTION

This project performs zero-shot topic classification using a pretrained language model.

The goal is to classify text samples into predefined topic categories without any model fine-tuning.

## 2. DATASET DESCRIPTION

The dataset consists of text entries sourced from a topics dataset typically containing

questions or statements belonging to categories such as Sports, Health, Family, Education,

Science, Business, Entertainment, and more. Each text sample is associated with a true topic label.

## 3. MODEL USED

A transformer-based zero-shot classification pipeline is used.

The model processes each text input along with a set of candidate topic labels and predicts

the most likely topic using natural language inference.

## 4. TRAINING SETUP

No fine-tuning was performed. The model operates entirely in zero-shot mode.

Inference was run on a batch of samples, recording both predicted labels and confidence scores.

## 5. EVALUATION METRICS

To evaluate performance, a confusion matrix and a heatmap were generated.

The confusion matrix shows correct predictions along the diagonal and misclassifications
off-diagonal.

General observations:

- Strong performance on distinct topics (e.g., Sports, Health).

- Overlapping topics (e.g., Science vs Education) show higher confusion.

- Overall accuracy estimated visually around 40–55%, typical for zero-shot classification.

## 6. CHALLENGES AND OBSERVATIONS

- Zero-shot models struggle with similar or semantically overlapping categories.

- Some classes with fewer examples show higher error rates.

- Zero-shot performance depends heavily on the phrasing of candidate labels.

- Larger descriptions for labels can improve accuracy.

## 7. CONCLUSION

The project demonstrates how zero-shot topic classification can be used without training data.

Although useful, fine-tuning or better label engineering could significantly improve accuracy.