أكاديمية طويق
Tuwaiq Academy

MULTI-CLASSIFICATION OF NEWS TOPICS

Presented by Reema Aldabass

أكاديمية طويق
Tuwaiq Academy

INTRODUCTION

This project focuses on building an AI-powered text classification system that automatically categorizes news articles into one of four categories: World, Business, Sports, or Science/Technology.

Using techniques from Natural Language Processing (NLP) and deep learning, the model reads the content of a news article and predicts its most likely category.

This helps automate the process of organizing large volumes of news data, making it faster and more efficient for users and media platforms.

أكاديمية طويق
Tuwaiq Academy

METHODOLOGY

01 Data Collection: Loaded AG News dataset from Hugging Face.

02 EDA: Explored data distribution and text lengths.

03 Text Cleaning: Removed punctuation, numbers, and converted text to lowercase.

04 Tokenization: Applied Byte Pair Encoding (BPE) for subword tokenization.

أكاديمية طويق
Tuwaiq Academy

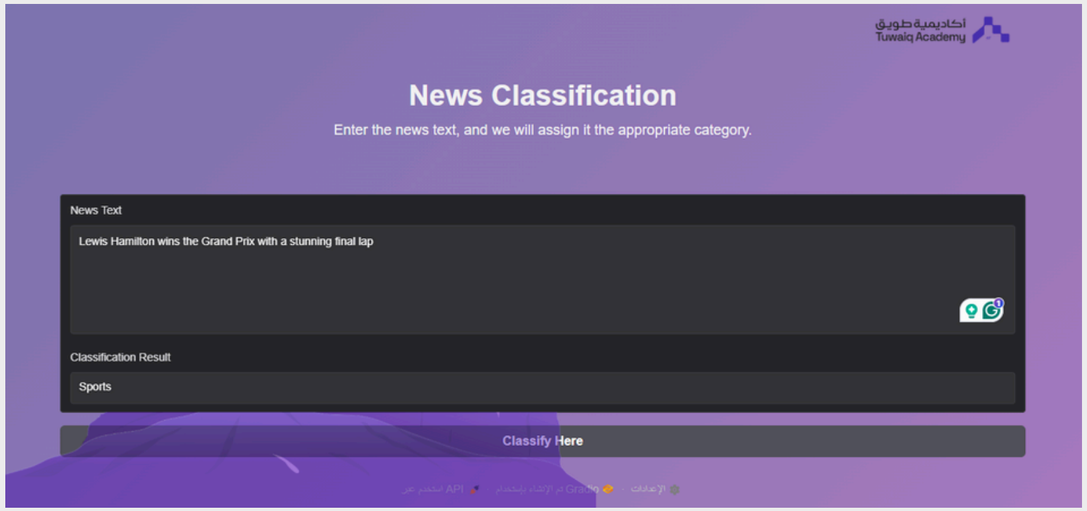
METHODOLOGY

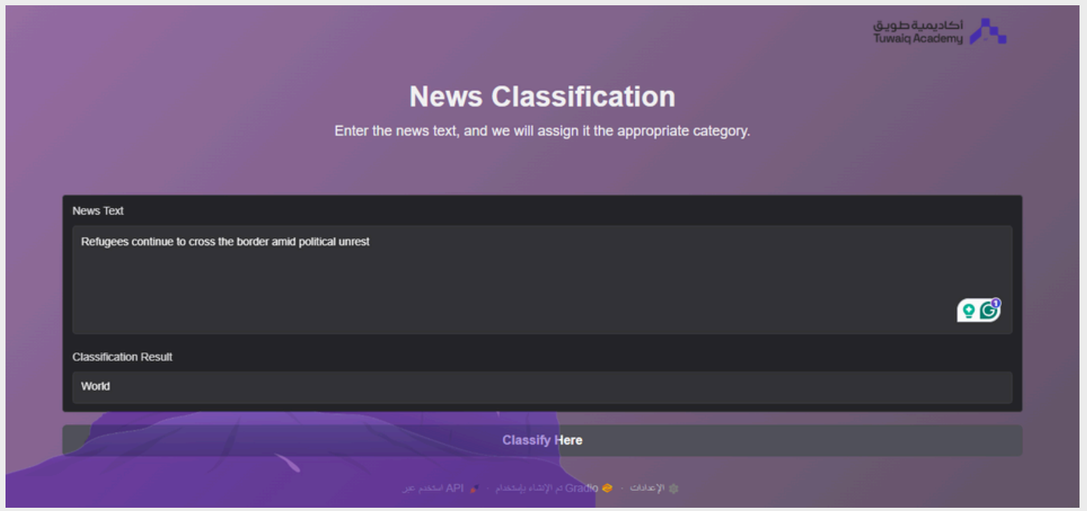
05 Numerical Encoding: Converted tokens into numerical IDs.

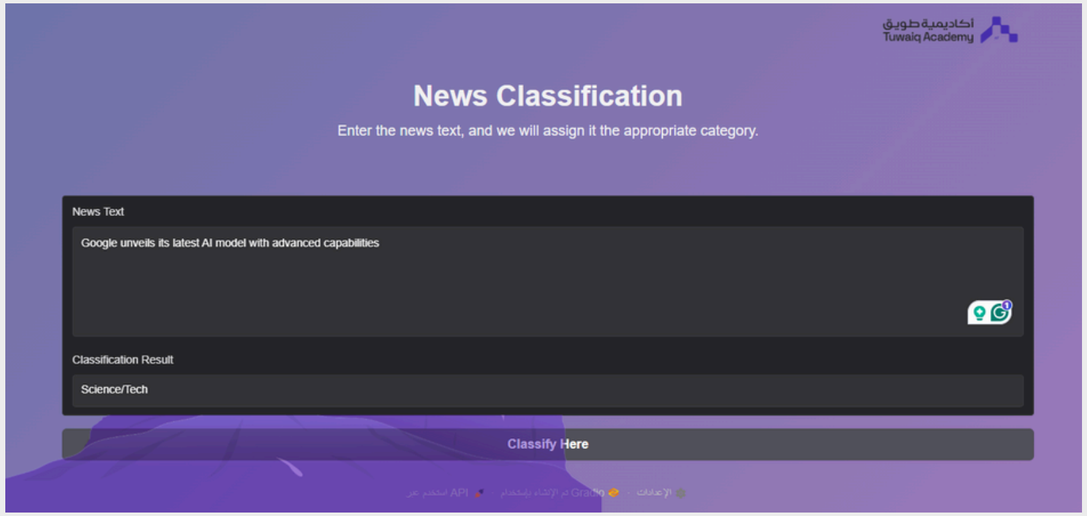
06 Model Building: Built an LSTM-based classification model.

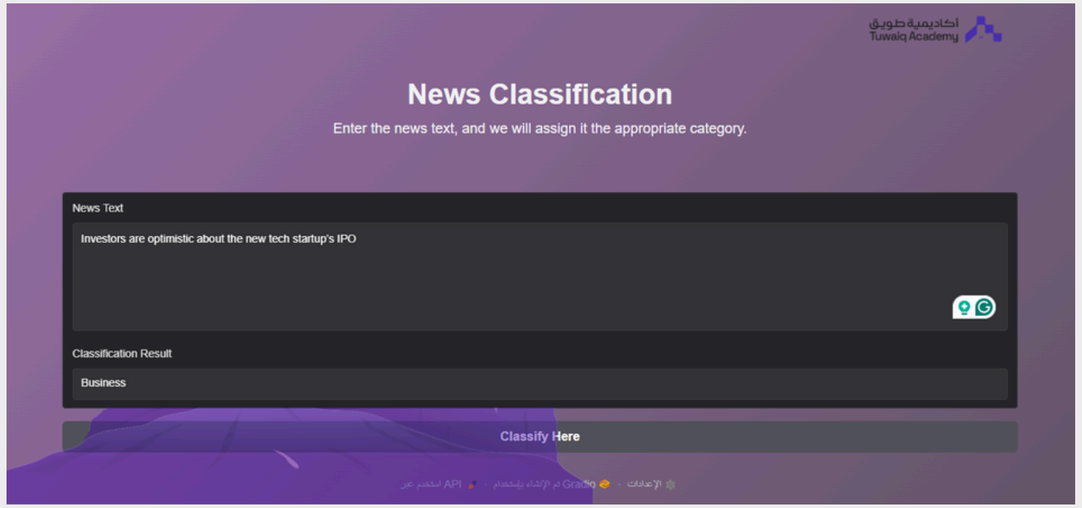
07 Gradio UI: Developed a web interface for real-time prediction.

IMPLEMENTATION









TOOLS USED

أكاديمية طويق
Tuwaiq Academy

PYTHON	The main programming language.
PYTORCH	For building and training the LSTM text classification model.
PANDAS	For data analysis and manipulation.
NUMPY	For numerical operations.
RE (REGULAR EXPRESSIONS)	For text cleaning and preprocessing.
GRADIO	To create an interactive web interface for the model.
MATPLOTLIB	For data visualization and plotting.

أكاديمية طويق
Tuwaiq Academy

THANK YOU

Presented by Reema Aldabass