News Articles Sorting

Architecture Document Design

Author: Tanjina Proma

Date: 1/08/2024

Contents

1. Introduction

1.1 Scope

The project aims to develop a robust text classification system tailored for news articles using BERT Base Uncased Transformers. Its scope involves implementing a model capable of accurately categorizing news content into relevant topics or classes. By harnessing BERT's contextual understanding, the system endeavors to capture nuanced relationships within articles for precise classification across diverse news domains.

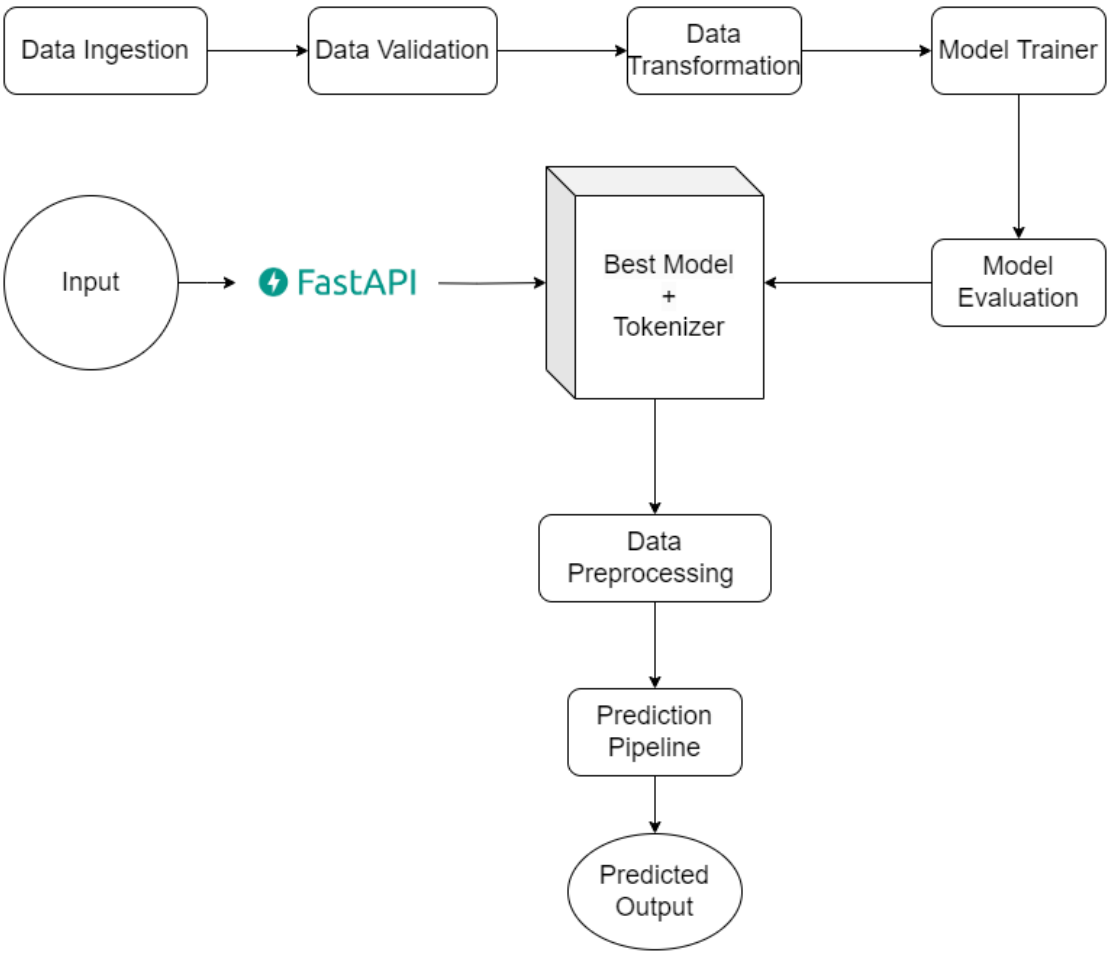
1.2 Objective

Objectives include fine-tuning the pre-trained BERT model on labeled news datasets, ensuring adaptability to varied article structures and themes. The system seeks to streamline the classification process, enabling efficient handling of large volumes of news data while maintaining high accuracy. Moreover, it aims to provide a user-friendly interface or API for seamless interaction, allowing users to submit articles for classification and retrieve categorized results promptly. Ultimately, the project strives to deliver a scalable, reliable, and accurate text classification solution specifically tailored for news articles, facilitating better content organization and analysis.

1.3 Significance of the project

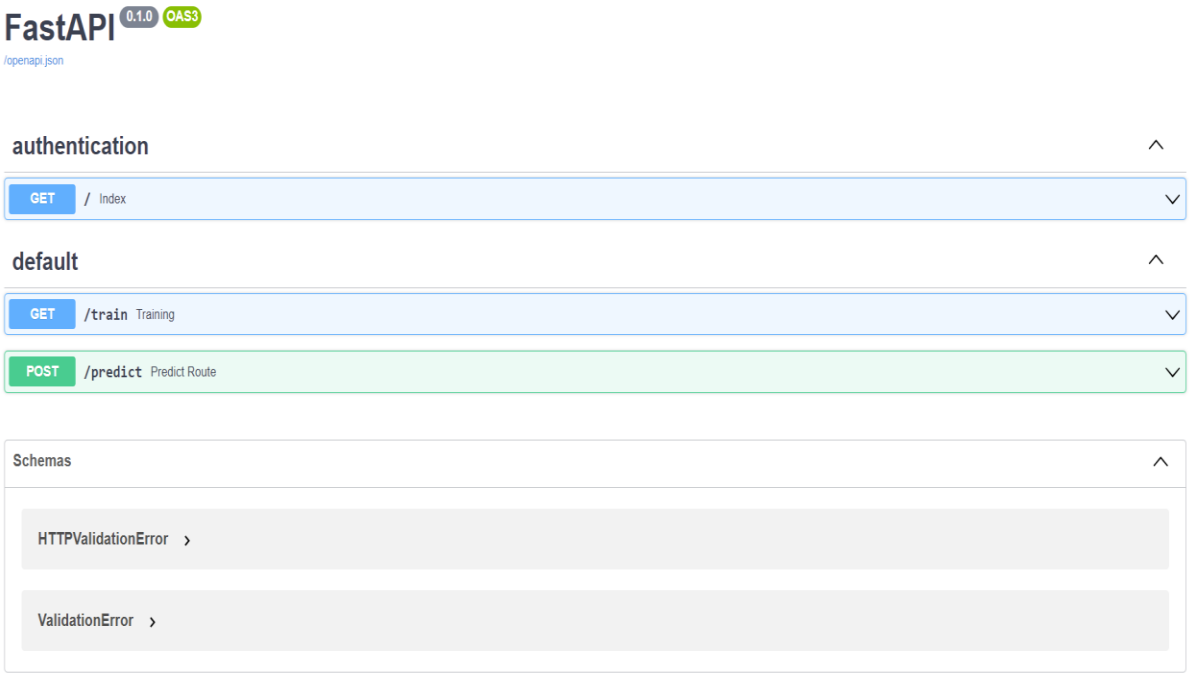
- **Information Organization:** Text classification structures vast amounts of news articles, aiding in organizing information based on topics, sentiments, or relevance.
- **Content Recommendation:** Enables personalized news delivery by understanding user preferences and recommending relevant articles.
- **Trend Analysis:** Identifies emerging topics or trends by categorizing articles, assisting in understanding public interests and sentiments.
- **Fake News Detection:** Helps in flagging or filtering out misinformation by classifying articles based on credibility or factuality.
- **Targeted Advertising:** Facilitates targeted advertising by classifying articles and understanding audience interests.
- **Search Optimization:** Enhances search functionalities by categorizing articles, improving search accuracy and relevance.
- **Policy Making and Governance:** Assists policymakers in understanding public sentiments, aiding in decision-making based on categorized news data.
- **Market Intelligence:** Helps businesses analyze market trends, customer sentiments, and competitor activities from news articles, guiding strategic decisions.

2. System Architecture Overview



The system architecture for text classification of news articles using BERT Base Uncased Transformers involves several key components. Raw news articles are ingested into the system, undergoing preprocessing before being tokenized for BERT input. The BERT Base Uncased model, known for its contextual understanding, processes these tokens, extracting high-level features and relationships. Fine-tuning the pre-trained BERT model enables task-specific classification. The output module generates classification predictions, assigning categories or labels to the articles. This architecture ensures the seamless flow of data from input through preprocessing, leveraging BERT for comprehensive contextual analysis, and finally producing accurate classifications. The robustness of BERT Base Uncased Transformers in capturing intricate linguistic nuances forms the backbone of this system, empowering precise categorization of news articles across diverse topics with enhanced comprehension and organization.

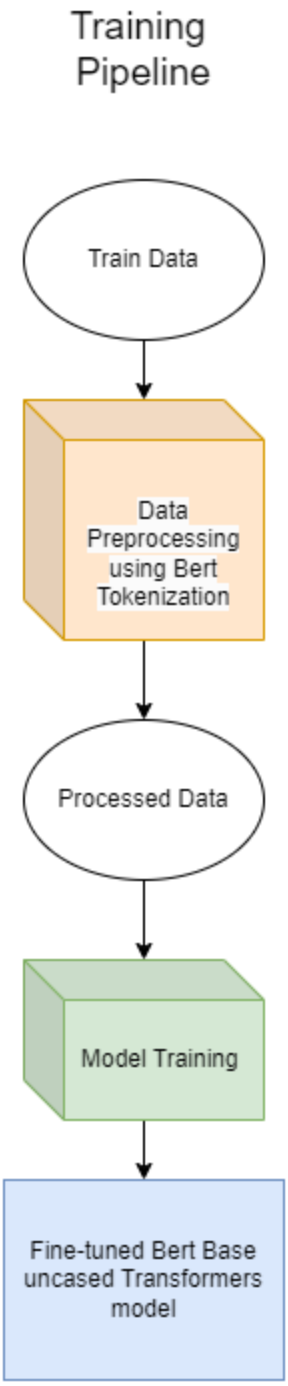
3. User I/O workflow

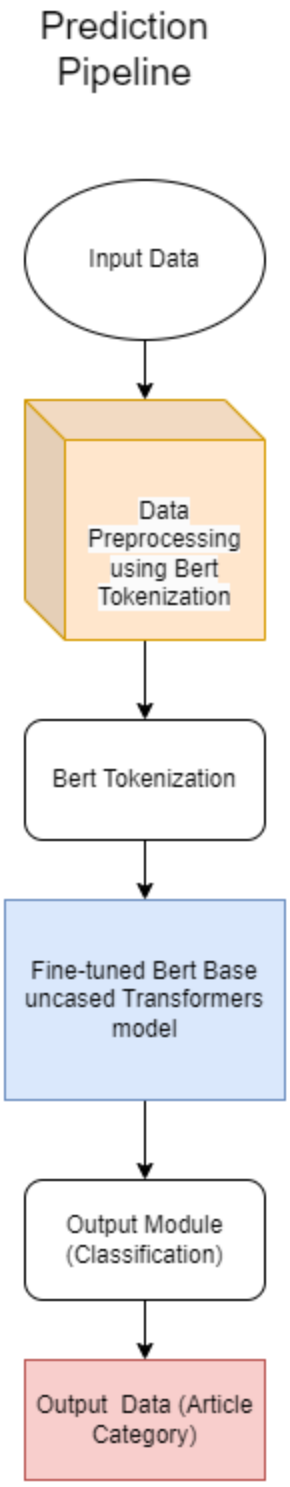


A FastAPI application facilitates the seamless transfer of data between users and a text classification model by acting as an interface that handles incoming requests, processes data, and delivers responses. When a user interacts with the FastAPI-based system, they send a request containing the necessary data, such as a text snippet for classification, via HTTP methods (e.g., POST or GET).

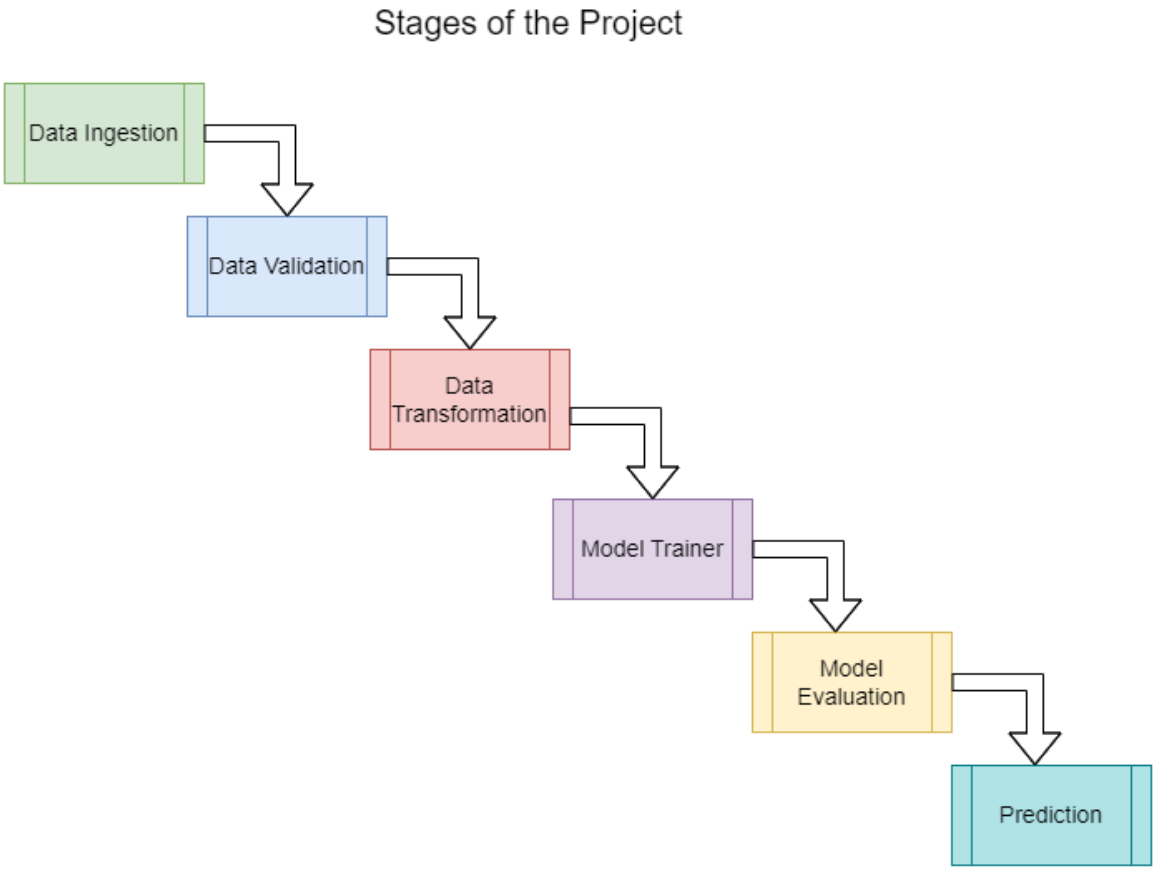
FastAPI's endpoints, defined with specific routes and functions, receive these requests, extracting the relevant information. The application preprocesses and validates the incoming data, preparing it for consumption by the text classification model. The prepared data is then passed to the model, which performs the classification task. Once the model predicts the category or label, the FastAPI application generates a response containing the classification results. This response is sent back to the user in a structured format (e.g., JSON), providing the classification output, completing the data transfer cycle between the user and the text classification model through the FastAPI application.

In this Application there are two pipelines available for the users to use. One is the training pipeline and the other is prediction pipeline.


Training Pipeline
Train Data
Data Preprocessing using Bert Tokenization
Processed Data
Model Training
Fine-tuned Bert Base uncased Transformers model


Prediction Pipeline
Input Data
Data Preprocessing using Bert Tokenization
Bert Tokenization
Fine-tuned Bert Base uncased Transformers model
Output Module (Classification)
Output Data (Article Category)

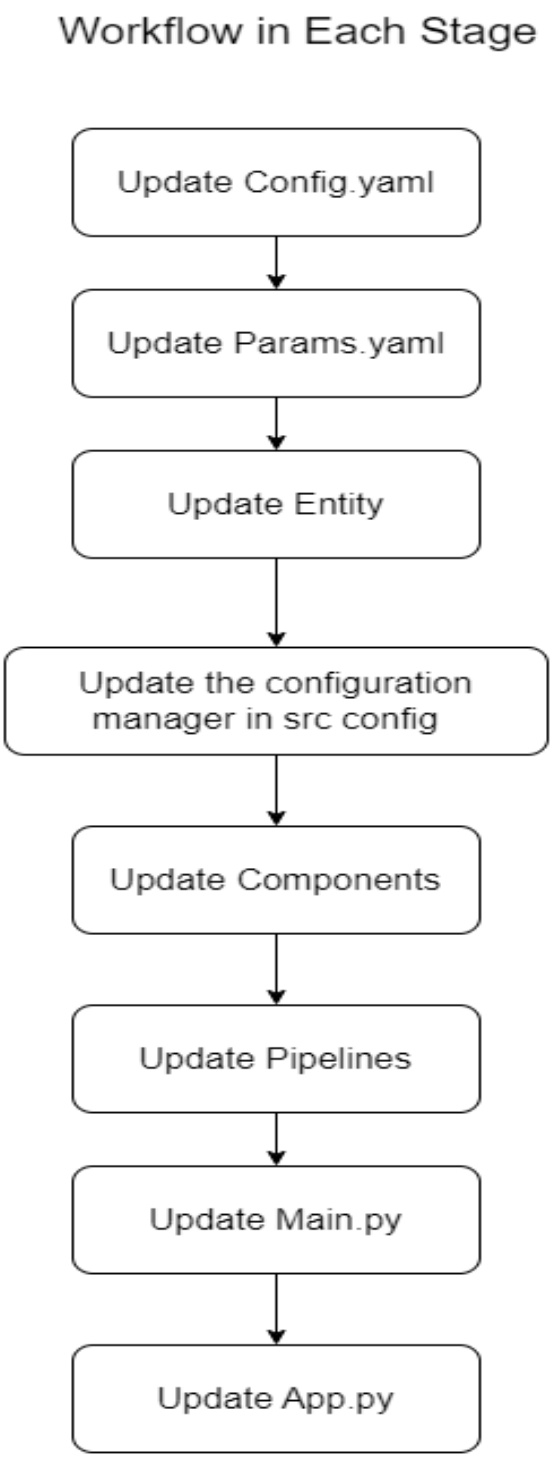
4. System Components



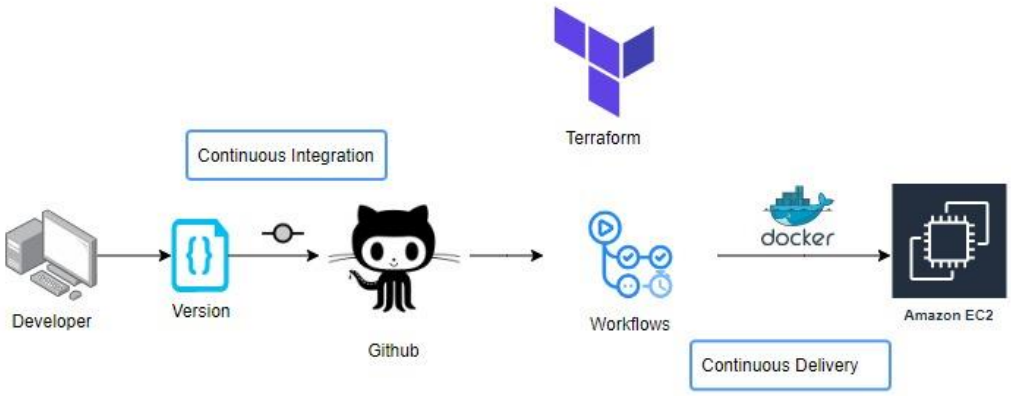
The text classification system for news articles using BERT Base Uncased Transformers comprises several key components. Firstly, the system relies on a data ingestion module to receive and preprocess raw news articles. The heart of the system lies in the BERT Base Uncased Transformer model, which leverages pre-trained language representations to extract intricate contextual embeddings from the articles.

An output module handles the classification predictions, assigning categories or labels to the articles based on the BERT model's outputs. Supporting these components are functionalities for data preprocessing, tokenization, and the fine-tuning phase of the BERT model on labeled news articles for task-specific classification. Additionally, the system includes mechanisms for model evaluation, ensuring accuracy, precision, recall, and F1-score metrics are measured to validate its performance. This architecture ensures a robust and comprehensive framework for accurately categorizing news articles by harnessing the power of BERT-based language understanding.

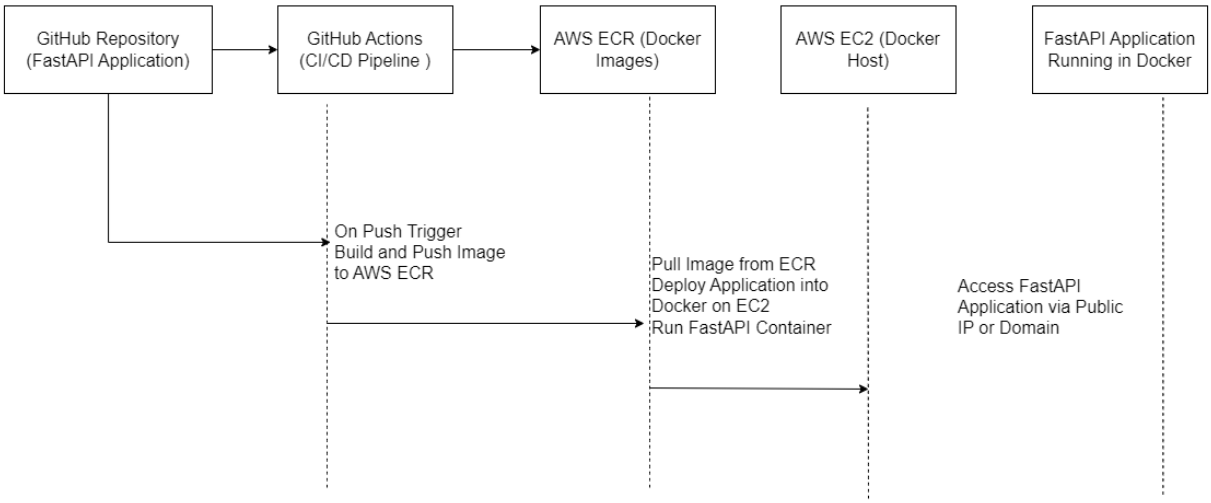
5. Components in Detail



6. Integration and Deployment



Wrokflow diagram for the Deployment Process



This workflow showcases the process starting from the GitHub repository, utilizing GitHub Actions for CI/CD to build and push Docker images to AWS ECR. Then, an EC2 instance pulls the Docker image from ECR and runs the FastAPI application within a Docker container. Finally, the application is accessible through the EC2 instance's public IP or domain.