**T5Bayes**

**Himanshu Dahiya**

**E22CSEU1144**

**Abstract:  
This project presents an integrated system combining T5-based text summarization and Naive Bayes sentiment analysis to process and analyze textual review data. Leveraging IMDb and product review datasets, the system preprocesses, trains, and saves models to generate concise summaries and predict sentiments efficiently. A user-friendly Streamlit app deploys the models, enabling real-time summarization and sentiment prediction from user-provided input or uploaded CSV files. By automating review analysis, this solution reduces manual effort, enhances usability, and provides actionable insights, making it valuable for domains reliant on customer feedback and large-scale text processing.**

**Introduction:  
Online reviews are pivotal for businesses and individuals seeking insights into user preferences and feedback. However, the sheer volume of reviews, often verbose and unstructured, poses significant challenges for manual analysis. Summarizing lengthy reviews and extracting sentiments are crucial tasks that simplify the decision-making process. This project aims to address these challenges by developing an automated system combining state-of-the-art text summarization and sentiment analysis techniques. The T5 transformer model is fine-tuned for abstraction-based summarization, while a Naive Bayes classifier is trained to predict review sentiments. The integration of these models into a Streamlit app provides users with an accessible and interactive platform for review analysis.**

**Proposed Work:  
The proposed work focuses on designing and implementing an automated pipeline for review analysis, comprising:**

1. **T5 Summarization Model: Fine-tuning the T5 transformer on a product review dataset to generate concise and contextually relevant summaries.**
2. **Naive Bayes Sentiment Analysis Model: Training a multinomial Naive Bayes classifier on the IMDb dataset to categorize reviews as positive or negative.**
3. **Streamlit App: Integrating both models into an interactive app that processes user input or CSV files to provide summaries and sentiment predictions in real-time.  
   This system is designed to be efficient, scalable, and adaptable for diverse domains.**

**Methodologies**:

1. **Data Collection and Preprocessing**:
   * **Datasets**: IMDb dataset for sentiment analysis and a product review dataset for summarization.
   * **Preprocessing**:
     + Clean text by removing stopwords, punctuation, and HTML tags.
     + Tokenize and prepare data for input into T5 and Naive Bayes models.
2. **Model Training**:
   * **T5 Summarizer**: Fine-tune the pretrained T5 model on the product review dataset to generate abstraction-based summaries.
   * **Naive Bayes Sentiment Analyzer**: Train a multinomial Naive Bayes model on the IMDb dataset to classify reviews as positive or negative.
3. **Integration**:
   * Save both models and integrate them into a Streamlit app.
   * Enable user input or CSV upload for review summarization and sentiment prediction.
4. **Deployment**:
   * Deploy the Streamlit app for interactive use.
   * Outputs include concise summaries and sentiment labels for each review.

**Experimental Results**:

* **T5 Summarizer**:
  + Fine-tuned model achieved high ROUGE and BLEU scores, indicating coherent and relevant summaries.
* **Naive Bayes Sentiment Analyzer**:
  + Demonstrated over 85% accuracy on IMDb test data, confirming reliable sentiment classification.
* **Integrated System**:
  + The Streamlit app effectively summarized and classified reviews, with user-friendly interfaces and real-time processing.

**Conclusion**:  
The integration of a T5 summarizer and Naive Bayes sentiment analyzer offers an efficient solution for automated review analysis. This system successfully combines advanced summarization capabilities with fast sentiment classification, reducing manual effort and enhancing decision-making. The deployed Streamlit app enables users to process reviews interactively, demonstrating the model's practical applicability.

**Future Scope**:

* **Multilingual Support**: Extend models to handle reviews in multiple languages.
* **Advanced Sentiment Analysis**: Upgrade the sentiment analyzer to a transformer-based model (e.g., BERT or Roberta) for nuanced detection.
* **Enhanced Deployment**: Deploy the solution as a web service or API for broader accessibility.
* **Custom Applications**: Adapt the system for domain-specific use cases, such as healthcare reviews or legal document analysis.