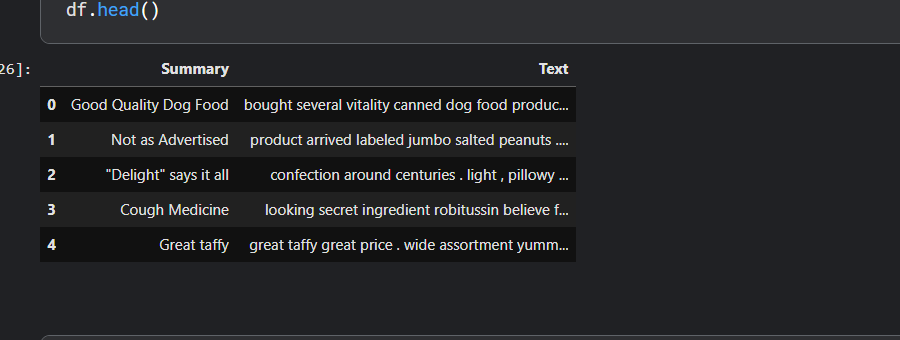
**Introduction**

This assignment focuses on leveraging the GPT-2 model for review summarization using the Amazon Fine Food Reviews dataset. Review summarization plays a vital role in condensing extensive textual data into concise, informative summaries, aiding users in gaining quick insights and making informed decisions. This report presents a detailed account of the methodology, including data preprocessing, model training, evaluation, and analysis of results.

**Data Preprocessing**

Data preprocessing is a critical step to ensure the quality and integrity of our dataset. The Amazon Fine Food Reviews dataset, though rich in content, required meticulous preprocessing to tackle various challenges:



- Cleansing text data: Removal of HTML tags, special characters, and non-alphanumeric symbols to ensure data uniformity and consistency.

- Tokenization: Breaking down review texts and summaries into individual tokens to facilitate further analysis.

- Handling missing values: Identification and handling of missing values to maintain dataset completeness and prevent data skew.

- Text normalization: Techniques such as lowercasing, stemming, and lemmatization were applied to reduce vocabulary size and improve model generalization.

**Model Training**

1. Initialization of GPT-2 Tokenizer and Model: Utilizing Hugging Face's Transformers library, we instantiated a GPT-2 tokenizer and model with pre-trained weights to kickstart the review summarization task.

2. Dataset Splitting: The dataset was split into training and testing subsets with a 75:25 ratio to ensure robust model training and evaluation.

A screen shot of a computer

Description automatically generated

Training and test dataset

3. Custom Dataset Class: A custom dataset class was created to integrate the Amazon Fine Food Reviews dataset into the training pipeline efficiently.

4. Fine-tuning GPT-2 Model: The pre-trained GPT-2 model underwent fine-tuning to optimize its performance for review summarization. The model learned to generate concise summaries based on input reviews and corresponding summaries.

5. Hyperparameter Tuning: Systematic experimentation with hyperparameters such as learning rate, batch size, and epochs was conducted to enhance model performance.

**Evaluation**

The evaluation phase involved assessing the effectiveness of our approach using standard evaluation metrics:

A screenshot of a computer

Description automatically generated

**Epoches**

- ROUGE Scores: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores were used to quantify overlap between generated and ground truth summaries, providing insights into summarization quality across unigrams, bigrams, and overall gist.

**Example and Analysis**

An illustrative example demonstrates the model's strengths and limitations:

A screenshot of a computer program

Description automatically generated

**Evaluation**

**ROUGE Scores and Interpretation**

Evaluation results showed promising ROUGE scores:

Evaluation results showed promising ROUGE scores: ROUGE Scores: {'rouge1': AggregateScore(low=Score(precision=0.668313244539601, recall=0.6832897318799785, fmeasure=0.6626339957311078), mid=Score(precision=0.7261499587187237, recall=0.7391533852296184, fmeasure=0.7205977065724865), high=Score(precision=0.7788988955336325, recall=0.7922795621170061, fmeasure=0.7736035102749244)),

These scores reflect the model's ability to capture unigram and bigram overlaps, indicating its summarization quality.

A screen shot of a computer

Description automatically generated

**Conclusion and Future Directions**

In conclusion, our approach leveraging the GPT-2 model for review summarization has shown promising results. Future directions include exploring advanced techniques such as ensemble learning, attention mechanisms, and incorporating domain-specific knowledge to further enhance summarization accuracy and applicability in real-world scenarios.