Executive Summary: Classification of Video Advertisements

1. Introduction

This project involves developing a classifier to answer 21 binary questions about video advertisements. The classifier uses both textual data extracted from the video and the provided textual descriptions. The aim is to maximize performance metrics (agreement percentage, F1 score, precision, and recall) compared to ground-truth data deduced from human coders.

2. Implementation

The implementation is divided into three main parts: analysis of ground truth data, processing video data, and merging the data for classification.

Part 1: Analysis of Ground Truth Data And Deduced Ground Truth Data(449 data to 150 data)

- **Data Loading and Inspection**: Ground truth and text data were loaded using pd.read_csv, and missing values were inspected with isnull().sum().
- **Dropping Irrelevant Columns**: Non-essential columns were identified and dropped using drop(columns=columns_to_drop, axis=1).
- **Deducing Majority Vote**: Functions majority_vote and majority_vote_special were defined to deduce the majority vote for binary and multi-class questions. The ground truth data was grouped by creative_data_id, and majority votes were calculated for each question.
- **Tokenization and Stopwords Identification**: Textual data was tokenized and stopwords were identified using NLTK functions word_tokenize and stopwords.words('english') in the clean_text function.

Part 2: Working with Video Data

- **Library Installation**: Necessary libraries for video processing and OCR were installed.
- **Text Cleaning**: The clean_text function was applied to clean textual data by removing punctuation, converting to lowercase, and removing stopwords.
- **OCR Function**: The function extract_frames_and_ocr was defined to extract frames from videos and apply OCR to extract text using cv2.VideoCapture and pytesseract.image_to_string.
- **Processing Videos**: Each video file was processed, the OCR function applied, and the results stored in a DataFrame.
- **Creating DataFrame**: A DataFrame video_data_df was created to store the extracted video text data.

Part 3: Merging Video and Textual Data

- **Merging Data**: The extracted video data was merged with the cleaned textual data on matching IDs using pd.merge.

- **Dropping Redundant Columns**: Redundant and original text columns were dropped using drop(columns=columns).
- **Combining Text Columns**: Text from video OCR, cleaned description, and cleaned speech was combined into a single column using apply(lambda row: ' '.join([...])).
- **Zero-Shot Classification**: Used a pre-trained zero-shot classification model to predict answers to predefined questions with the pipeline function from the transformers library.
- **Prediction Extraction and Evaluation**: Predictions were converted to a DataFrame, multi-class responses were mapped to binary values, and F1 score, precision, recall, and accuracy were calculated for 20 questions using `f1_score`, `precision_score`, `recall_score`, and `accuracy_score` from `sklearn.metrics`. And for one of the question: "Is there any verbal or visual mention of the price?", was handled differently by mapping positive responses ('Yes, visual'; 'Yes, both') to 1 and 'No' to 0 to align with the binary responses in `predictions_df`

3. Evaluation Metrics

The classifier was evaluated based on agreement percentage, F1 score, precision, and recall. Below are the F1 scores for each question:

Question	F1 Score	Precision	Recall	Accuracy
Is there a call to go online (e.g., shop online, visit the Web)?	0.36	0.27	0.53	0.46
Is there online contact information provided (e.g., URL, website)?	0.58	0.48	0.74	0.52
Is there a visual or verbal call to purchase (e.g., buy now, order now)?	0.50	0.36	0.84	0.39
Does the ad portray a sense of urgency to act (e.g., buy before sales ends, order before ends)?	0.39	0.29	0.57	0.52
Is there an incentive to buy (e.g., a discount, a coupon, a sale or "limited time offer")?	0.57	0.45	0.77	0.49
Is there offline contact information provided (e.g., phone, mail, store location)?	0.28	0.17	0.70	0.33
Is there mention of something free?	0.17	0.09	0.92	0.29
Does the ad mention at least one specific product or service (e.g., model, type, item)?	0.88	0.85	0.90	0.78
Does the ad show the brand (logo, brand name) or trademark multiple times?	0.91	0.84	0.98	0.83
Does the ad show the brand or trademark exactly once at the end of the ad?	0.90	0.89	0.90	0.82
Is the ad intended to affect the viewer emotionally?	0.89	0.87	0.91	0.80
Does the ad give you a positive feeling about the brand?	0.79	0.86	0.74	0.67

Does the ad have a story arc, with a beginning and an end?	0.40	0.29	0.68	0.49
Does the ad have a reversal of fortune?	0.23	0.13	0.95	0.16
Does the ad have relatable characters?	0.63	0.52	0.79	0.53
Is the ad creative/clever?	0.74	0.65	0.85	0.61
Is the ad intended to be funny?	0.33	0.20	0.94	0.22
Does this ad provide sensory stimulation?	0.64	0.58	0.71	0.54
Is the ad visually pleasing?	0.71	0.73	0.69	0.59
Does the ad have cute elements like animals, babies, animated characters, etc.?	0.40	0.28	0.71	0.51

4. Documentation

The process and results were documented thoroughly:

- **Methodology**: Described each step in the data processing and classification pipeline.
- **Results**: Included a detailed table of F1 scores, precision, recall, and accuracy for each question.
- **CSV File**: Submitted a CSV file with the final answers per video per question.

5. Submission

The complete code and the final CSV file were submitted as required.

6. Bonus Questions

Analysis of Classifier Performance

Certain videos might not work well with the classifier or yield inconsistent answers for several reasons:

1. **Quality of OCR**: Low-quality video frames or OCR errors can lead to inaccurate text extraction, affecting classification performance.
2. **Ambiguity in Content**: Videos that mix multiple messages or lack clear indicators for certain questions can confuse the classifier.
3. **Complex Visual and Auditory Cues**: Ads with complex visual or auditory cues that are not easily captured by text analysis may result in inconsistent answers.
4. **Generalization Issues**: The zero-shot classification model might struggle to generalize across different advertisement styles and content, leading to variability in performance.

In-Depth Analysis of Human Coders' Responses and Classifier Performance

1. **Human Coders' Variability**:

    - Human coders' interpretations can vary based on subjective understanding, leading to inconsistencies in ground-truth data.
    - Ensuring consistent training and guidelines for human coders can help mitigate this issue.
2. **Classifier Performance**:
    - The classifier generally performs well on questions with clear textual indicators, such as brand mentions and calls to action.
    - Questions involving emotional impact or creativity, which are more subjective, show more variability in performance.

Observed Patterns and Anomalies

1. **High Agreement on Objective Questions**:
    - Questions related to specific product mentions, brand logos, and trademarks have high agreement percentages and F1 scores, indicating clear indicators in the content.
2. **Low Agreement on Subjective Questions**:
    - Questions involving emotional response, humor, and creativity have lower agreement percentages, reflecting the subjective nature of these questions and the challenge for both human coders and the classifier to consistently interpret them.
3. **Potential Causes**:
    - **Content Complexity**: Ads with complex narratives or mixed messages can cause both human coders and classifiers to struggle with consistent interpretation.
    - **Quality of Data**: Inaccurate OCR results due to poor video quality can affect the classifier's ability to correctly interpret the content.

Conclusion

The developed classifier demonstrates reasonable performance across various questions, with room for improvement in handling ambiguous or complex video content. Future work could focus on enhancing OCR accuracy, refining text analysis techniques, and incorporating additional data features to boost classification performance. The analysis highlights the importance of clear and objective indicators for accurate classification and the challenges posed by subjective interpretation of advertisement content.