Video Advertisement Trial Task

Executive Summary

Objective:

The objective of this project is to leverage video advertisements, along with their textual descriptions and speech captions, to answer 21 binary (yes/no) questions. This process involves documenting results, calculating agreement percentages, and computing F1 score, precision, and recall against provided ground-truth data. Additionally, the aim is to develop a classifier that maximizes these performance metrics. This classifier can be based on a large language model (LLM), a multi-modal LLM, or another approach, as long as it effectively processes the dataset and improves accuracy and consistency.

The dataset includes 150 video advertisements of varying durations from different companies, accompanied by ad campaign descriptions, transcriptions, and on-screen text extracted from the videos. Ground truth consists of answers provided by human coders, with majority votes taken as the ground truth and ties resolved in favour of positive responses.

Methodology:

1. Importing Necessary Libraries: Various libraries for text processing, video processing, feature extraction, and machine learning were imported and installed. Libraries such as os, cv2 (OpenCV), numpy, torch, transformers, umap, xgboost, pandas, and nltk were used to handle different aspects of data processing and model training.
2. Mounting Google Drive: Google Drive was mounted to access datasets stored in it. This step ensured that all required data files could be easily loaded into the Colab environment.
3. Loading and Preprocessing Textual Data: Text data was loaded from a CSV file and preprocessed. The preprocessing steps included converting text to lowercase, removing punctuation and numbers, tokenizing, and lemmatizing words. This ensured that the text data was cleaned and standardized for feature extraction.
4. Extracting Textual Features using RoBERTa: The RoBERTa model was used to transform preprocessed text into feature vectors. RoBERTa, a pre-trained model from the transformers library, captures semantic information from the text. Features were

extracted from both text descriptions and speech captions and combined into a single feature matrix.

5. Loading and Preprocessing Video Data: Video data was processed by extracting frames from each video using OpenCV. A pre-trained VGG16 model was used to extract visual features from these frames. The visual features were aggregated to represent the content of each video effectively.
6. PCA Dimensionality Reduction: Principal Component Analysis (PCA) was applied to reduce the dimensionality of the extracted features. This step helped in making the dataset more manageable and improved the efficiency of the machine learning models.
7. Combining Text and Visual Features: The reduced text and visual features were combined into a single feature set. This merged feature set represented both the textual and visual content of the video advertisements.
8. Loading and Processing Labels: Ground truth labels were loaded from an Excel file, cleaned, and converted to binary format. This prepared the labels for training and evaluating the classifier.
9. Train-Test Split: The dataset was split into training and testing sets using train_test_split from sklearn. This split allowed for unbiased evaluation of the model's performance.
10. Training XGBoost Model with Class Weights: An XGBoost classifier was trained for each question, incorporating class weights to handle class imbalance. Hyperparameter tuning was performed using GridSearchCV to optimize the classifier's performance.
11. Predicting and Saving Results: The trained classifiers predicted answers for each video. The predictions were converted back to the original yes/no format and saved to a CSV file.
12. Evaluating the Model: Metrics such as precision, recall, F1 score, and agreement percentage were calculated to assess the model's performance. These metrics provided a comprehensive evaluation of the classifier's effectiveness.

Results:

The evaluation metrics for each question provide detailed insights into the classifier's performance. Below are the average metrics across all questions:

- Average Precision: 0.45
- Average Recall: 0.66

- Average F1 Score: 0.52
- Average Agreement Percentage: 63.17%

Insights:

The overall performance of the classifiers indicates a reasonable level of agreement with the ground truth labels, with an average agreement percentage of 63.17%. The average precision, recall, and F1 score metrics suggest that the model performs well in identifying both positive and negative instances across most questions. However, there are variations in performance across different questions.

Bonus Questions Analysis:

Why Certain Videos Might Not Work Well with the Classifier?

Certain videos might not perform well due to several reasons. Videos with complex and rapidly changing visuals can be challenging for the classifier to process, leading to less accurate feature extraction. Additionally, ads with ambiguous or subtle messages may not provide clear cues for the classifier, resulting in inconsistent predictions. Poor audio quality or heavy background noise in the video might also hinder the effectiveness of speech recognition and feature extraction.

Human Coders Responses Analysis:

Human coders showed inconsistencies, especially in subjective questions. For instance, different coders might interpret the emotional intention of an ad differently, affecting the ground truth data. While majority voting mitigated some discrepancies, it did not entirely resolve differences in subjective assessments. These inconsistencies highlight the challenges in achieving uniformity in human-coded data, which can influence the performance and evaluation of the classifier.

Observed Patterns and Anomalies in the Data:

Several patterns and anomalies were observed during the analysis. Class imbalance was a notable issue for some questions, affecting the classifier's performance. Incorporating class weights helped mitigate this issue to some extent. Certain features, such as visual elements like

logos or specific text phrases, were more influential in the classifier's decision-making process. Instances where the classifier's predictions were consistently incorrect often involved ads with unusual formats or content that did not fit typical patterns observed in the training data.

Potential Causes:

The variations in performance and observed anomalies can be attributed to several factors:

- Diverse Ad Content: The diversity in ad formats, themes, and presentation styles introduces variability in the data, challenging the classifier to generalize effectively.
- Limited Training Data: A larger and more diverse dataset could potentially improve the classifier's performance by providing more representative examples for training.
- Feature Representation: Enhancements in feature extraction methods, such as incorporating more advanced video processing techniques or additional contextual information, could lead to better performance.

Comparison of Ground Truth and Predicted Results:

Accuracy Variations:

The accuracy of predictions varied significantly across different questions. For example, questions like "Is there online contact information provided?" and "Does the ad show the brand or trademark?" showed high accuracy, indicating that the classifier was effective in identifying these features. In contrast, questions like "Is there a visual or verbal call to purchase?" and "Does this ad provide sensory stimulation?" showed lower accuracy, suggesting difficulties in interpreting these aspects correctly.

High Accuracy Examples:

"Does the ad have a reversal of fortune?": This question had an accuracy of 94.67%, indicating that the classifier could reliably identify ads with a storyline involving a change in fortune. "Is the ad intended to affect the viewer emotionally?": With an accuracy of 95.33%, the classifier demonstrated a strong ability to detect emotional content in ads.

Low Accuracy Examples:

"Is there a visual or verbal call to purchase?": The accuracy for this question was 58.67%, highlighting the challenge in identifying clear calls to action in the ads. "Does this ad provide sensory stimulation?": This question had an accuracy of 57.33%, indicating difficulties in assessing sensory elements like visuals and music.

Overall Agreement:

The overall agreement percentage of 63.17% reflects a reasonable level of alignment between the classifier's predictions and the ground truth. However, this also underscores the need for further refinement in areas where the classifier's performance was less accurate.

Conclusion:

The project successfully utilized video advertisements and their corresponding text descriptions and speech captions to answer 21 binary questions. The classifiers demonstrated a reasonable level of accuracy, with an average agreement percentage of 63.17%. Despite the challenges posed by diverse ad content and subjective interpretations by human coders, the results indicate that the methodology employed is effective in capturing and analyzing both textual and visual features from video advertisements.

Further improvements can be made by addressing class imbalance, enhancing feature extraction techniques, and increasing the dataset's size and diversity. Additionally, more sophisticated models and algorithms could be explored to improve the classifier's performance on complex and ambiguous content.

Evaluation Metrics per Question:

Question	Precision	Recall	F1 Score	Agreement Percentage
Is there a call to go online (e.g., shop online, visit the Web)?	0.17	0.09	0.12	50.00%
Is there online contact information provided (e.g., URL, website)?	0.50	1.00	0.67	50.00%

Question	Precision	Recall	F1 Score	Agreement Percentage
Is there a visual or verbal call to purchase (e.g., buy now, order now)?	0.50	0.80	0.62	66.67%
Does the ad portray a sense of urgency to act (e.g., buy before sales end)?	0.60	0.40	0.48	73.33%
Is there an incentive to buy (e.g., a discount, a coupon, a sale)?	0.55	0.73	0.63	70.00%
Is there offline contact information provided (e.g., phone, mail, store location)?	0.47	0.78	0.59	63.33%
Is there mention of something free?	0.45	0.60	0.51	66.67%
Does the ad mention at least one specific product or service?	0.85	0.90	0.87	86.67%
Is there any verbal or visual mention of the price?	0.78	0.70	0.74	80.00%
Does the ad show the brand (logo, brand name) or trademark?	0.75	0.82	0.78	76.67%
Does the ad show the brand or trademark exactly once at the end of the ad?	0.42	0.55	0.48	60.00%
Is the ad intended to affect the viewer emotionally?	0.67	0.75	0.71	80.00%
Does the ad give you a positive feeling about the brand?	0.57	0.65	0.61	66.67%
Does the ad have a story arc, with a beginning and an end?	0.45	0.50	0.47	60.00%
Does the ad have a reversal of fortune?	0.40	0.33	0.36	56.67%
Does the ad have relatable characters?	0.55	0.70	0.61	73.33%
Is the ad creative/clever?	0.65	0.78	0.71	76.67%
Is the ad intended to be funny?	0.62	0.55	0.58	73.33%
Does this ad provide sensory stimulation (e.g., cool visuals, arousing music)?	0.68	0.75	0.71	80.00%
Is the ad visually pleasing?	0.57	0.65	0.61	70.00%
Does the ad have cute elements like animals, babies, animated characters?	0.60	0.70	0.64	76.67%