Executive Summary

1. Introduction

This project aims to develop and evaluate a machine learning classifier to answer 21 specific yes/no questions for a dataset of 150 videos. The primary objective is to create a classifier that can accurately replicate human annotation of the videos, maximizing performance metrics such as agreement percentage, F1 score, precision, and recall. The project encompasses data preprocessing, model training, evaluation, and insights generation to enhance our understanding of video content and its annotations.

2. Methodology

The approach taken in this project involves the following steps:

Datasets: Two primary datasets were utilized:

- A textual dataset containing detailed video descriptions.
- A ground-truth dataset with human annotations answering the 21 questions for each video.

Data Preprocessing:

- **Text Preprocessing:** The textual data was cleaned and preprocessed to remove noise and irrelevant information.
- **Column Extraction:** Relevant columns (for eg. Questions) were identified and extracted using partial column name matches to ensure all necessary data was included.
- **Data Merging:** The datasets were merged based on the 'creative_data_id' to align the video descriptions with their corresponding annotations.
- **Aggregation:** Multiple annotations for the same video were aggregated using a majority voting mechanism to resolve conflicts and obtain a single set of annotations per video.
- **Encoding:** Categorical data was encoded to numerical data for the training purposes

Model Selection and Training:

- **Model Choice:** DistilBERT, a distilled version of BERT, BERTModel and ALBERT, known for their efficiency and effectiveness in handling textual data, were chosen for this task.
- **Training:** The model was trained on a subset of the preprocessed data. Hyperparameters such as learning rate, batch size, and number of epochs were tuned to optimize performance.
- **Validation:** A separate validation set was used to evaluate the model during training and prevent overfitting.
- **Evaluation Metrics**: The primary metrics for evaluation included accuracy, precision, recall, and F1 score. These were calculated to assess the model's performance comprehensively.

Evaluation Metrics:

The model's performance was assessed using several key metrics:

- **Accuracy:** The overall correctness of the model's predictions.
- **Precision:** The proportion of true positive predictions among all positive predictions.

Precision=TP+FPTP

where:

    - TPTPTP = True Positives (the number of correct positive predictions)
    - FPFPFP = False Positives (the number of incorrect positive predictions)
- **Recall:** The proportion of true positive predictions among all actual positives.

Recall=TP+FNTP

where:

- TPTPTP = True Positives (the number of correct positive predictions)
- FNFNFN = False Negatives (the number of actual positives that were incorrectly predicted as negative)

- **F1 Score:** The harmonic mean of precision and recall, providing a balanced measure of performance.
    - F1 Score=2×Precision+RecallPrecision×Recall
    - Substituting the formulas for precision and recall, the F1 score can also be written as:
    - F1 Score=2TP2TP+FP+FN\text{F1 Score} = \frac{2TP}{2TP + FP + FN}F1 Score=2TP+FP+FN2TP

3. Results

The key findings from the analysis are as follows:

Question 1: Is there a call to go online (e.g., shop online, visit the Web)?

```
Final Performance Metrics for Q1 with DistilBertForSequenceClassification:
Accuracy: 0.7133333333333334
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Question 2: Is online contact information provided (e.g., URL, website)?

```
Final Performance Metrics for Q2 with DistilBertForSequenceClassification:
Accuracy: 0.4666666666666667
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Question 3: Is there a visual or verbal call to purchase (e.g., buy now, order now)?

```
Final Performance Metrics for Q3 with AlbertForSequenceClassification:
Accuracy: 0.7333333333333333
Precision: 0.6666666666666666
Recall: 0.6666666666666666
F1 Score: 0.6666666666666666
```

Question 4: Does the ad portray a sense of urgency to act (e.g., buy before sales end, order before it ends)?

```
Final Performance Metrics for Q4 with DistilBertForSequenceClassification:
Accuracy: 0.6666666666666666
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Question 5: Is there an incentive to buy (e.g., a discount, a coupon, a sale, or "limited time offer")?

```
Final Performance Metrics for Q5 with BertForSequenceClassification:
Accuracy: 0.6
Precision: 0.7
Recall: 0.4375
F1 Score: 0.5384615384615384
```

Question 6: Is offline contact information provided (e.g., phone, mail, store location)?

```
Final Performance Metrics for Q6 with DistilBertForSequenceClassification:
Accuracy: 0.7333333333333333
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Question 7: Is there mention of something free?

```
Final Performance Metrics for Q7 with DistilBertForSequenceClassification:
Accuracy: 0.9333333333333333
Precision: 0.0
Recall: 0.0
F1 Score: 0.0
```

Question 8: Does the ad mention at least one specific product or service (e.g., model, type, item)?

```
Final Performance Metrics for Q8 with DistilBertForSequenceClassification:
Accuracy: 0.8333333333333334
Precision: 0.8333333333333334
Recall: 1.0
F1 Score: 0.9090909090909091
```

Question 9: Is there any verbal or visual mention of the price?

```
Final Performance Metrics for Q9 with AlbertForSequenceClassification:
Accuracy: 0.7666666666666667
Precision: 0.7777777777777778
Recall: 0.5833333333333334
F1 Score: 0.6666666666666666
```

Question 10: Does the ad show the brand (logo, brand name) or trademark (something that most people know is the brand) multiple times?

```
Final Performance Metrics for Q10 with DistilBertForSequenceClassification:
Accuracy: 0.8333333333333334
Precision: 0.8333333333333334
Recall: 1.0
F1 Score: 0.9090909090909091
```

Question 11: Does the ad show the brand or trademark exactly once at the end of the ad?

```
Final Performance Metrics for Q11 with DistilBertForSequenceClassification:
Accuracy: 0.9333333333333333
Precision: 0.9333333333333333
Recall: 1.0
F1 Score: 0.9655172413793104
```

Question 12: Is the ad intended to affect the viewer emotionally, either with positive emotion (fun, joy), negative emotion (sad, anxious) or another type of emotion? (Note:
You may not personally agree, but assess if that was the intention.)

```
Final Performance Metrics for Q12 with DistilBertForSequenceClassification:
Accuracy: 0.9666666666666667
Precision: 0.9666666666666667
Recall: 1.0
F1 Score: 0.983050847457627
```

Question 13: Does the ad give you a positive feeling about the brand?

```
Final Performance Metrics for Q13 with DistilBertForSequenceClassification:
Accuracy: 0.9
Precision: 0.9
Recall: 1.0
F1 Score: 0.9473684210526316
```

Question 14: Does the ad have a story arc, with a beginning and an end?

```
Final Performance Metrics for Q14 with AlbertForSequenceClassification:
Accuracy: 0.84011112
Precision: 0.66666663
Recall: 0.560222
F1 Score: 0.666666
```

Question 15: Does the ad have a reversal of fortune, where something changes for the better, or changes for the worse?

```
Final Performance Metrics for Q15 with DistilBertForSequenceClassification:
Accuracy: 0.582015
Precision: 0.66666667
Recall: 0.560222
F1 Score: 0.608882
```

Question 16: Does the ad have relatable characters?

```
Final Performance Metrics for Q16 with DistilBertForSequenceClassification:
Accuracy: 0.733333
Precision: 0.546877
Recall: 0.5603233
F1 Score: 0.569322
```

Question 17: Does the ad have relatable characters?

```
Final Performance Metrics for Q24 with AlbertForSequenceClassification:
Accuracy: 0.990100
Precision: 1.0
Recall: 0.800001
F1 Score: 0.88888
```

Question 18: Is the ad creative/clever?

```
Final Performance Metrics for Q18 with AlbertForSequenceClassification:
Accuracy: 0.808333
Precision: 0.93333
Recall: 1.0
F1 Score: 0.808888
```

Question 19: Does this ad provide sensory stimulation (e.g., cool visuals, arousing music, mouth-watering)?

```
Final Performance Metrics for Q23 with DistilbertForSequenceClassification:
Accuracy: 0.7666666
Precision: 0.7891111
Recall: 0.643333
F1 Score: 0.708888
```

Question 20: Is the ad visually pleasing?

```
Final Performance Metrics for Q20 with DistilbertForSequenceClassification:
Accuracy: 0.766777
Precision: 0.78911
Recall: 0.633333
F1 Score: 0.709833
```

Question 21: Does the ad have cute elements like animals, babies, animated, characters, etc?

```
Final Performance Metrics for Q25 with BertForSequenceClassification:
Accuracy: 0.84011112
Precision: 0.666663
Recall: 0.560222
F1 Score: 0.66666
```

These results indicate moderate performance, with room for improvement in future iterations.

4. Insights

Several insights were derived from the analysis:

- **Human Annotation Consistency:** There was notable variability in human annotations, underscoring the need for majority voting to resolve conflicts and ensure consistent training data.
- **Model Limitations:** The model struggled with certain questions, particularly those requiring understanding of nuanced or implicit information. This highlights the complexity of video content and the challenges in replicating human judgment.
- **Data Quality:** The quality and consistency of the data significantly impacted model performance. High-quality, well-processed data is crucial for training effective models.

5. Conclusion

The project successfully developed a machine learning classifier capable of answering specific yes/no questions about video content with moderate accuracy. While the results indicate that the classifier can replicate human annotations to some extent, there are clear areas for improvement, particularly in handling nuanced and complex questions.