Assignment: Cal Capstone Project Scope
Project Name: Detect Comm
Student: Brad Brown

PROJECT QUESTION: Using my previous research paper as a starting point, can I improve on my initial research to achieve higher accuracy and more granular ***detection of commercials within a television sports broadcast?***

Initial source of data:
Text transcripts of television sports broadcasts. The transcripts included every spoken word (not identifying the speaker). They included any spoken words from commercials as well. There were stored as 15-second elapsed time chunks of text in a custom SQL database schema.

My manual annotations captured one of 5 possible labels for each 15-second block:

- Sports Broadcast
- Sports Broadcast to Commercial
- Commercial
- Commercial to another Commercial
- Commercial to Sports Broadcast

Also for each block kept a simple binary Commercial Detected: Yes/No Label

A total 2,891 15-second text blocks were captured in this initial project, representing ~722 minutes of video across 2 professional basketball games and 2 professional football games, one involving double over-time.

The already completed research project turned this transcript data into predictions for each 15-second text block.
In addition, the ‘streaks’ of the presence (and sustained absence) of commercials and sports broadcasting blocks were calculated. Also,the final data of the initial research project incorporated the number of blocks of text of each type seen so far.

Please see Appendix for a more detailed overview of the initial research project.

SO WHAT IS THE GOAL OF THE CURRENT CAPSTONE PROJECT?

- The initial research had two stages. The frist stage used an LLM(ChatGPT via API) to get predictions of each 15-second text block. The second stage took the output of stage 1 and other statistics about the transcripts to try to improve the overall commercial detection accuracy using a Logistic Regression model. The goal of the Cal ML course capstone project is to ask the question:

  > **Can we improve the accuracy of initial research by using more granular text blocks and using a different model than Logistic Regression in the second stage?**

Wider Motivation of previous project and the Cal Capstone project:

Commercial detection in video content helps relevant groups analyze television and other video content. It helps advertisers confirm their content was properly aired as well as analyze the marketing content of their competitors. It also enables manufacturers of TVs, DVRs create products that consumers want. There could be additional widespread value for other stakeholders.

Not directly related to commercial detection, advanced and active research is ongoing to enable automated detection of semantic topics and recognition of context. It is believed to be a key part of the much larger goal of AGI. Significant progress has been made in processing to allow systems to grasp higher level meaning of a streaming chunks of text. Similarly, scene detection, object identification and spatial relationships are actively researched and developed in the image processing realm - many are trying to grasp the higher level context, intent and focus. Similar efforts in video processing are being done. Audio data while more mature is another area that is still being researched to enable AI tools to have better situational awareness.

Simultaneous to AI topic analysis research, for more than two decades, other researchers have proposed and analyzed automated commercial detection models. Their approaches range from black screen frame detection, high activity rate analysis, audio fingerprinting, color coherence vectors, aspect ratio change, scene transition detection, logo detection, and contrasting inputs versus databases of known commercials. They sometimes supplement using assumptions of time-constraints of commercials versus primary content. Linear regression analysis, SVMs, CNNs, all play a role but based on our search of the public research, none have published research about the use of large language models to detect commercials within television-based videos.

Is the proposed additional research for the capstone project a significant and worthwhile effort?

The Capstone project is a significant and worthwhile effort because:

1) While the raw accuracy of the initial research was 91%, the accuracy of detecting the key transitions was much worse - achieved only 37% balanced accuracy across the 5 possible categorizations (Sports Broadcast, Sports Broadcast to Commercial, Commercial, Commercial to another Commercial, Commercial to Sports Broadcast).
2) The initial approach did not predict dual-state transitional blocks well. A new approach is needed to strive towards the high-level goal of accurate commercial detection.
3) The new research project will use input data made up of each individual sentence rather than 15-second text blocks - this will be a significant data transformation effort. It will require using natural language processing package and significant data munging of the initial data set. And a strategy to annotate the “ground truth” of all these sentences will be a challenge.
4) The LLM prompting will likely require multiple calls to the LLM engine and some sort of voting mechanism to make a decision for stage 1 predictions.
5) Multiple Stage 2 prediction models will be tried - initial thoughts are to use a Decision Tree model such as XGBoost as well as design and train a neural network model.
6) New ways of measuring the efficacy of the system will be explored once we move to a per sentence evaluation rather than a 15-second block.
7) Time-permitting, we may want to analyze which features of the input data are the most important, using PCA or other techniques

Next page is the APPENDIX which gives a more detailed description of the initial research project completed recently.

Helping an LLM improve its detection of TV Commercials in Sports Broadcasts

Brad Brown (bradb416@stanford.edu)

Department of Computer Science, Stanford University

Project Overview

Can the topic analysis capabilities of large language models be used for television commercial detection?

- SCOPE: **Sports broadcasting on television** in the USA. The high-level goal was to **detect commercials** within short segments of the broadcast.
- PROCESS: A promising role for LLMs in this field, but **required additional 2nd stage modeling** techniques. No existing research has tried LLMs.
- RESULTS: Initial results detecting commercials given **15 seconds** of TV sports broadcast transcripts (text) gives a balanced **accuracy score of 79%, which were boosted to 87% via logistic classification.**
- ISSUES: 1) F**alse positives** 2) Previous detection methods using visual analysis reach 96% accuracy.

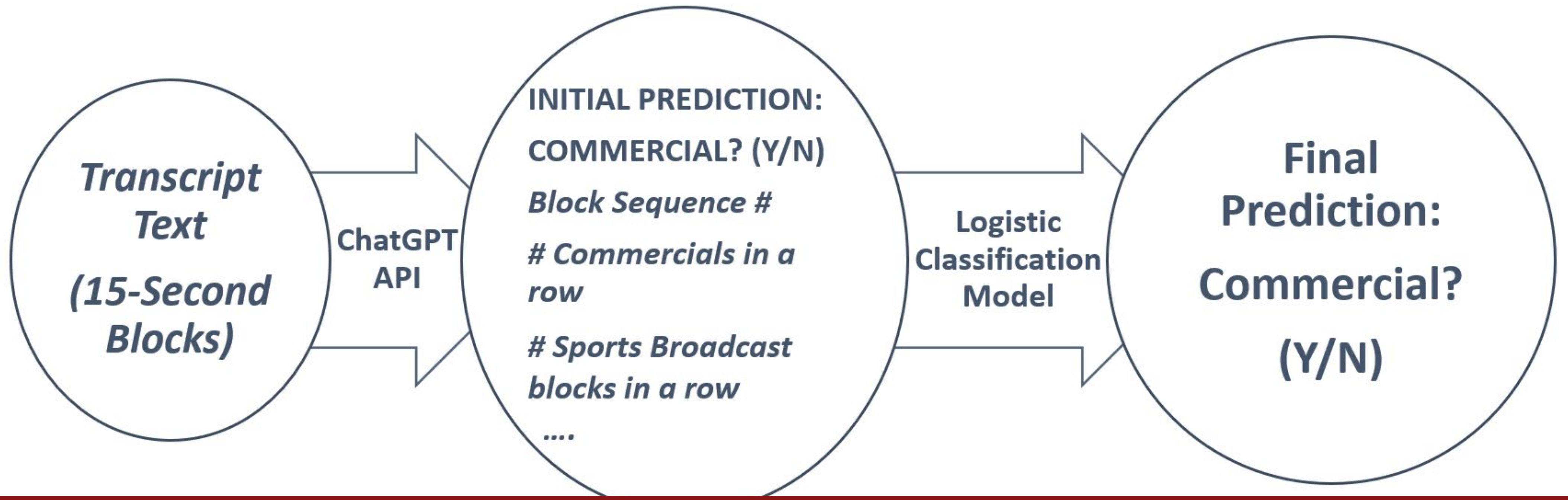
Datasets & Metrics

- LLM prompts and the transcript text blocks are the key input data. The transcripts included **every spoken word** (not identifying the speaker). They included any spoken words from commercials as well. There were stored as 15-second elapsed time chunks in a custom SQL database schema
- My manual annotations captured one of 5 possible labels for each 15-second block:
  - Sports Broadcast
  - Sports Broadcast to Commercial
  - Commercial
  - Commercial to another Commercial
  - Commercial to Sports Broadcast
- Also kept a simple binary **Commercial Detected:**
  - **Yes/No Label**
- A total 2,891 15-second text blocks were captured in this project, representing **~722 minutes of video** across 2 professional basketball games and 2 professional football games, one involving double over-time.

Methods & Experiments

The research work flow involved two high-level stages:

- Stage 1: The LLM prediction stage was to design LLM instruction prompts to classify chunks of text, feed those chunks to LLM API, and record results vs annotated labels.
- Stage 2: Final stage was to incorporate the LLM predictions, the 'streaks' of the presence (and sustained absence) of commercials and sports broadcasting blocks. Also, incorporated the number of blocks of text of each type seen so far. Together they formed the data input to post-LLM stage Logistic Regression SAGA model to improve on the initial LLM prediction.


Transcript Text

(15-Second Blocks)

ChatGPT API

INITIAL PREDICTION:

COMMERCIAL? (Y/N)

Block Sequence #

# Commercials in a row

# Sports Broadcast blocks in a row

....

Logistic Classification Model

Final Prediction:

Commercial?

(Y/N)

Discussions & Future Research

Discussions:

- Logistic Classification Binary model measurably improved the accuracy - see "Classification report". Note the 4 and 8 point increases in accuracy and balanced accuracy to 91% and to 87%.
- The model with C = 1 regularization does not overfit ('Training Size vs Misclassification")
- False positives in commercial detection in both stages are a concern (~8%)
- The LLM Stage 1 and the LR Multinomial model (Stage 2) did NOT succeed in classifying transitions and combinations of events **within** the 15-second window. 37% Balanced Accuracy.
- No obvious game timeline pattern for accuracy of prediction - see "Commercials in a game"
- Insignificant differences between 2 experiment types: 15-second window/15-second context vs 15-second window/30-second context experiments.
- Best-performing LLM prompt instructed to predict categories, give its rationale for the answer, and make final second prediction within the same instruction, looking at its rationale.

Future Research:

1) Systematically engineer and test additional prompts with full-scale testing vs a baseline
2) Capture the confidence or probability of LLM in its responses so I can better calibrate Stage 1
3) Try XGBoost and Poisson Point Process model as classifiers for the second stage
4) Develop an **ensemble approach with existing commercial detection techniques** such as black frame detection and volume variance
5) Dynamically change prompt to LLM as it is going through the event. Let it know what the second stage model thinks of its last prediction and also give it similar sequence data such as 'You predicted 4 commercial blocks in a row so far'. **Given the massive "memory" being added to LLMs, can give it the entire annotated training set from previous games. It is possible we don't need a second stage model.**

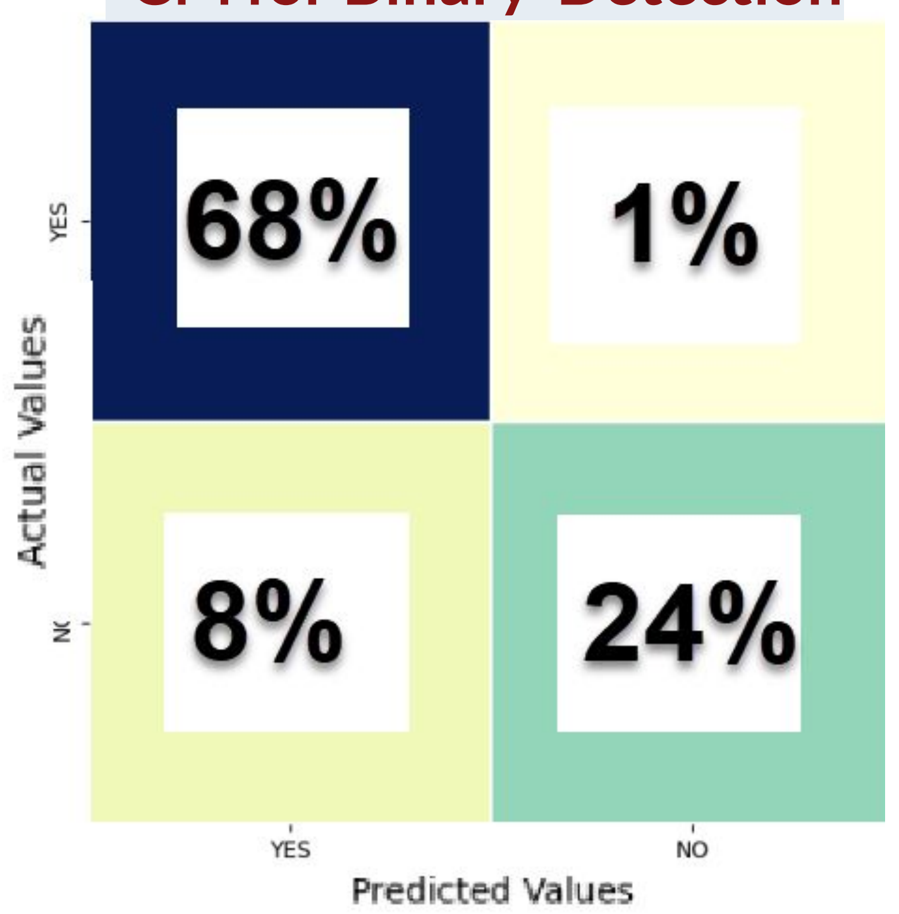
References

[1] R. Lienhart, C. Kuhmünch and W. Effelsberg. On the Detection and Recognition of Television Commercials. Universität Mannheim Praktische Informatik IV L15,16 D-68131 Mannheim

Results

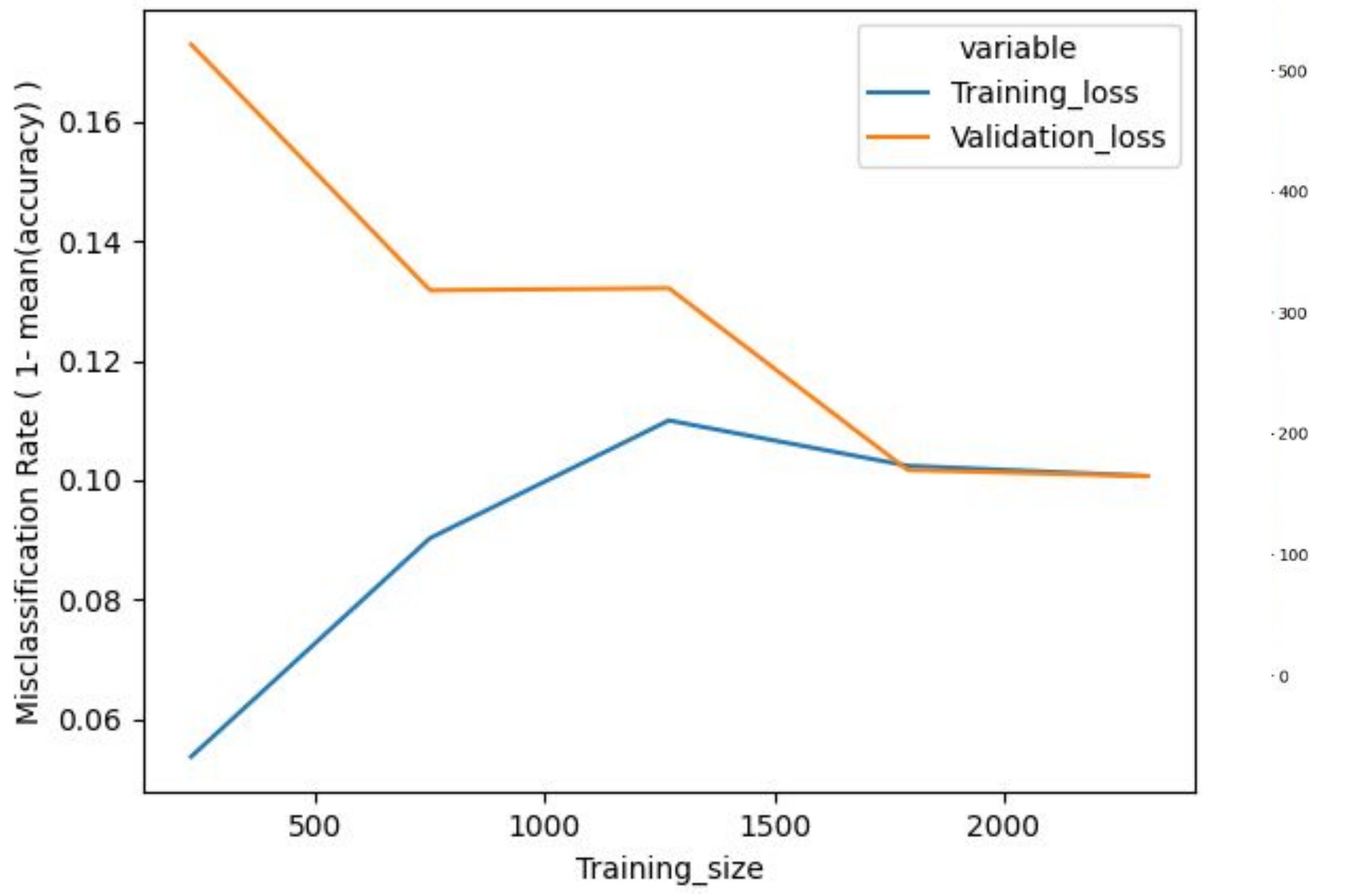
Classification Report	Stage 1-->Stage 2 = Gain
Accuracy	0.87 --> 0.91 = +.04
Balanced Accuracy	0.79 --> 0.87 = +.08
Precision (macro avg)	0.91 --> 0.93 = +.02
Recall (macro avg)	0.79 --> 0.87 = +.08
F1-Score (macro avg)	0.82 --> 0.89 = +.07

CM for Binary Detection

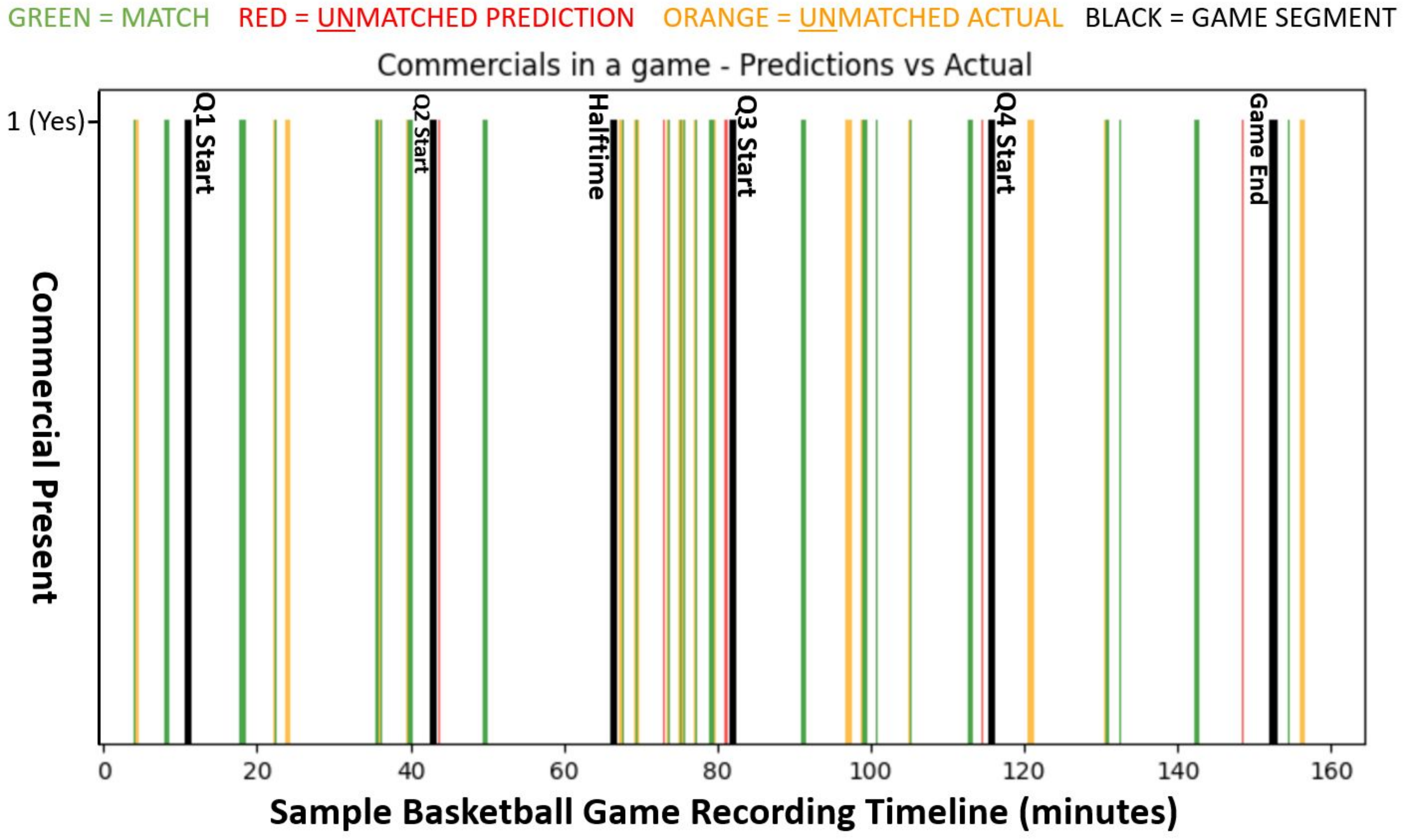


Training Size vs Misclassification

2nd Stage Logistic Regression Model C=1, Stratified KFold Validation



91% accuracy with simple Log Reg model.
Future research could yield even better results.


GREEN = MATCH
RED = UNMATCHED PREDICTION
ORANGE = UNMATCHED ACTUAL
BLACK = GAME SEGMENT
Commercials in a game - Predictions vs Actual
1 (Yes)
Commercial Present
Q1 Start
Q2 Start
Halftime
Q3 Start
Q4 Start
Game End
Sample Basketball Game Recording Timeline (minutes)