COMPSCI 4NL3: Natural Language Processing
Team Proposal

Team 4

Junnan Li
Nawaal Fatima
Rashad Bhuiyan
Sumanya Gulati

24 January 2025

Contents

List of Tables

List of Figures

1 Task Overview –Item 2

2 Task Definition –Item 3

3 Data Source and Plan for Data Collection –Item 4

Our data will be interview transcripts on asapsports.com, specifically the responses given by the interviewees (players and/or coaches). We will write an automated web scraping script in Python to collect the transcript text.

There is neither a terms of service nor a Robots Exclusion Protocol (robots.txt) file on the website, likely because the website doesn't get a lot of traffic. Since this is publicly available data taken from interviews, the biggest concern would be rate limiting. We would only need to scrape about 100 transcripts to meet our required number of data points, thus we can just limit the number of requests per second to not overload their servers.

After tokenizing and removing unecessary information (e.g. interviewee name) we can store the corpus in a single file or multiple files for labelling purposes.

4 Dataset Details –Item 5

A data point will be considered a single paragraph of a response to an interview question by a player or coach. For each event, we will look at each and every day of interview recordings that are tracked. The events covered this time will be the entirety of both NBA and WNBA Finals, as well as both NBA and WNBA Drafts. Based on these events, there are a total of 17 days of interviews. Furthermore, each recording day has interviews from multiple different players and coaches that each answer questions with either a one or multi-paragraph response. Based on this information, our dataset is expected to contain approximately 2500 data points.

The following is a set of 3 data points taken from the Game 2 Postgame interview with Jason Kidd of the Dallas Mavericks:

1. Yeah, we are not down. We're positive. This is a group that believes. We didn't get an opportunity to get a split or win two here on the road. Now Boston held serve. Now we've got to go home and hold serve.
    - **Assigned Labels: Positive, Team**
2. Big. The small things, you know, we have to do the small things, and that's part of the game. Those are points that we left on the board, and we didn't shoot free throws well tonight, and we have to be better.
    - **Assigned Labels: Negative, Team**
3. Yeah, I think Luka is a special player. He's one of, if not the best player in the world, and he causes a problem. He's able to find guys. Again, creating open opportunities, and we just didn't take advantage of it.
    - **Assigned Labels: Positive, Individual**

5 Team Contract –Item 6

5.1 Team Purpose

5.2 Team Member Roles

5.3 Facilitation Activities

5.4 Anything Else???