IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 February 2024, accepted 14 May 2024, date of publication 20 May 2024, date of current version 24 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3403490

APPLIED RESEARCH

Social Networks and Large Language Models for Division I Basketball Game Winner Prediction

GINA SPRINT, (Senior Member, IEEE)
Department of Computer Science, Gonzaga University, Spokane, WA 99258, USA
e-mail: sprint@gonzaga.edu

ABSTRACT Sporting event outcome prediction is a well-established and actively researched domain, with a particular focus on college basketball's March Madness tournament. Researchers, fans, and gamblers alike seek accurate game-level predictions using features such as tournament seeds, season performance, and expert opinions. While machine learning algorithms have been harnessed to build prediction models, no perfect model or human-created bracket has emerged. This paper explores a novel approach to basketball game outcome prediction by utilizing the power of social networks and large language models (LLMs). LLMs are trained to understand and generate text, often eliminating the need for a feature engineering step. Consequently, our method utilizes tweets from official Division I college basketball team Twitter accounts in the days leading up to a game as context for knowledge discovery and winner prediction with LLMs. To do this, we have compiled a comprehensive dataset of over one million tweets from both men's and women's teams spanning two consecutive seasons. Instead of relying on traditional numeric features, we employ only tweet text with few-shot/zero-shot learning, thereby offering an emerging social network-based approach for sporting event outcome prediction. Furthermore, using chain of thought prompting we investigate the information in team tweets that are predictive of future game performance.

INDEX TERMS Few-shot/zero-shot learning, in-context learning, large language models, March Madness prediction, NCAA college basketball, social networks.

I. INTRODUCTION

Sporting event outcome prediction is a long-standing and highly active research area spanning multiple disciplines. When it comes to college basketball and more specifically March Madness, researchers, fans, and gamblers alike want to make the best game-level predictions possible. While many prediction approaches exist, people often use tournament seeds, regular season matchup results, expert opinions, and their own intuition [1], [2], [3], [4], [5], [6], [7]. More specifically, statisticians and data scientists typically use large amounts of numeric data, like historical team performance, custom ordinal rankings, and player information, to train machine learning algorithms to try and build accurate prediction models. These techniques range from classical algorithms like k-nearest neighbors, logistic regression, support vector machines, random forests, and neural networks [7], to more recent methods like tree boosting and deep learning. Since 2014, Kaggle, the world's largest data science community, has hosted an annual March Machine Learning Mania competition [3], [4]. In 2023, this competition offered a $50,000 award to the team or individual submitting the best predictions as measured by the highest Brier score on the competition leaderboard [8]. While many submissions are based on predictions from deep neural networks, recent winning submissions have used XGBoost [9]. Even with incredible amounts of data, expert feature engineering, and unprecedented compute power, no perfect prediction model or human-created bracket has been created [10].

A simple, yet reasonably well performing approach to creating a bracket or predicting a game winner is to always choose the team with the higher seed/rank [6]. This baseline technique correctly predicts tournament game winners about 71% of the time [11]. Considering the best machine learning models typically yield accuracies in the range of about 67% to 75% [2], [7], [12], higher seed-based prediction offers a

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Magno.

reasonable benchmark for comparison. Of the 20+ million brackets submitted for the 2023 men's tournament, the highest ranked bracket correctly predicted 50 out of 63 games, which is a remarkable 79% accuracy [13]. Considering the variability from year-to-year, these numbers suggest that about a quarter of all tournament games will be difficult-to-predict upsets, with men's basketball typically exhibiting more upsets than women's basketball. Clearly, predicting basketball game winners is still an open research problem with opportunities for innovative solutions.

Recent advancements in natural language processing and large language models (LLMs) have redefined what machine learning is capable of in terms of solving new problems and revisiting the solutions of existing problems. LLMs are large transformer models trained to "understand" language by being fed large amounts of text from various sources during pre-training. LLM abilities have recently skyrocketed in depth and breadth due to several breakthroughs, most notably in terms of computational power, number of model parameters, and training dataset size [14]. These advances have been most noticeable and accessible via open-ended chat interfaces like OpenAI's ChatGPT. Models like those in the GPT family are generally used to accept text input in the form of a prompt (often with context about a task) and produce a text output in the form of a completion (hopefully performing the task). With text input, this format eliminates the need for a numeric feature engineering step, allowing text to be used directly for a wide variety of learning tasks [15]. Furthermore, in-context learning techniques, such as few-shot learning where examples of a task are provided [16], can prompt an LLM to perform diverse tasks without explicit training.

Since previous March Madness prediction approaches have focused on using numeric input data, we were interested in leveraging timely social network text and LLMs for an entirely new approach to basketball game winner prediction. To do this, we used recent tweets from each team's official Twitter (recently rebranded as X) account as context to prompt an LLM to predict which team will win the upcoming game. For few-shot learning, we included demonstrations of recent tweets/game outcomes. LLMs are particularly well suited for this task because predictive information is embedded in team tweets in diverse and latent forms. Subsequently, this work offers a new perspective on how timely social network activity, LLMs, and in-context learning can impact the field of sports analytics. To enable future work in this area, the dataset and code are available in our Github repository.[1]

II. RELATED WORK

Due to recent breakthroughs in LLM development and deployment, researchers in nearly all fields are exploring using text directly as input for prediction tasks. Recent studies have shown that LLMs can be used as a high performing approach for text-based understanding, generation, and prediction in a variety of domains [15], such as education [17], medicine [18], energy consumption [19], human mobility [19], [20], and stock prediction [21]. With this work, we expand the scope of LLMs to the sports analytics domain. With our new, large-scale social network dataset and LLM classification approach, we present a novel attempt at game outcome prediction.

While our approach to game prediction work is first-of-its-kind, research using social media for sports analytics has been investigated for over a decade [22]. Twitter has historically allowed liberal access to its massive database via its application programming interface (API), allowing researchers to create big datasets and investigate a wide variety of topics. In sports, most of these studies analyzed tweets from fans [23] or players [24], [25] for sentiment analysis [26] and/or social network analysis [27]. Far fewer studies have focused on collecting and analyzing tweets from official sport teams at the collegiate level [28], [29], [30]. One example is the work by Chen *et. al* which collected data from the official websites, Facebook pages, and Twitter accounts of Southeastern Conference institutions [29]. The goal of the work was to compare the men's and women's team coverage and usage of these communication platforms. Though not Twitter-based, another relevant study is the work of Wallace et al. that analyzed how institutions in the Big 12 Conference used Facebook for fan interaction and brand management [30].

Most similar to the present study, some investigations have used crowd opinions on social media to predict sporting event outcomes [26], [31], [32], [33]. UzZaman *et. al* used Twitter data to make game outcome predictions for the 2010 FIFA World Cup, motivating their approach by stating: "The main idea of using social media to predict the game is to let people do all these analysis and write about it in the social media, so that we can aggregate their predictions to make a guess on the outcome of the game." The authors created a system called TwitterPaul that parsed tweets and extracted predictions, achieving a precision of up to 88% [31]. Schumaker et al. also explored using tweets to predict game outcomes; however, their approach focused on using tweet sentiment to predict wins and spread in the English Premier League [26]. The authors found positive tweet sentiment to be predictive of match outcomes and point spread, possibly yielding higher betting payouts than traditional metrics like odds-favorites. In conclusion, the authors suggested there is predictive power in the hidden information contained within tweets. Beal *et. al* also predicted match outcomes for the English Premier League; however, they used machine learning with both statistical match data and text from sports journals to achieve a 63.2% accuracy, representing a 6.9% improvement over traditional methods with match statistics [34].

To advance social network and sports analytics research, our approach exhibits several differences from the previously described studies. First, our collegiate dataset is more comprehensive. We used social media data from all Division I men's and women's basketball teams from all conferences and across multiple seasons. Second, we explored the

[1] https://github.com/gsprint23/DivisionIBasketballTwitter

predictive power in the hidden information of official team tweets instead of fan tweets. If such predictive power exists, a deployed prediction system needs to only monitor two social media accounts instead of streaming tweets/posts from several stakeholder accounts and/or all relevant hashtags. In same-day betting scenarios, these just-in-time predictions could augment traditional game outcome prediction approaches (like the approach of Beal *et. al*) to increase prediction accuracy and/or predict an upset.

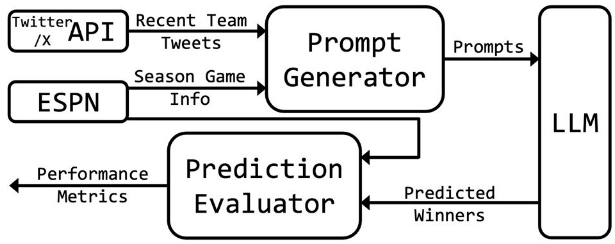


FIGURE 1. Overview of LLM-based game winner prediction.

III. METHODS

Figure 1 provides an overview of our proposed LLM-based game winner prediction approach. To undertake the approach, we first collected social media text and game matchup information. Though text from any social media platform could be utilized, we used Twitter to collect tweets created by each team due to its popularity and generous API access[2] for academic research that was available at the time. To collect the most recent season schedules and game outcome information for each team, we used the ESPN website. We collected this data for the most recent two college basketball seasons (2021-22 and 2022-23) to observe year-to-year social network trends and to provide a sufficient dataset for our LLM-based prediction approaches. These approaches were evaluated using the 2023 men's and women's tournament games.

A. DATA COLLECTION AND PROCESSING

To build a multi-season college basketball Twitter dataset, we first identified the current Division I institutions from the NCAA institution directory website.[3] For the 2022-23 season, this list included 363 institutions, including one new institution starting its reclassification in 2021 (University of St. Thomas) and five new institutions starting their reclassifications in 2022 (Lindenwood University, Queens University of Charlotte, University of Southern Indiana, Stonehill College, Texas A&M University – Commerce). Of these 363 institutions, 2 did not have women's basketball teams (The Citadel and Virginia Military Institute). For each Division I institution, we searched for their official men's and women's basketball team Twitter handles (also known as usernames), compiling a list of 724 official team Twitter handles. Using the Twitter API `/2/users/by` endpoint, we were able to lookup the unique Twitter user ID for each handle. While a Twitter user may change their handle, they cannot change their account's user ID, though they can create a new account. Henceforth in our analysis, we used these user IDs to represent the team accounts to avoid any ambiguities should a team change their handle. On the men's side, all but four of the accounts were created prior to the 2021-22 season. On the women's side, this number was five. In these cases, the team may have had a previous Twitter account that was deleted or suspended, causing the team to move to a new account (and hence receive a new user ID). The oldest account on the men's side was Duke University, created on 12/20/2008. On the women's side, the oldest account was Vanderbilt University, created on 2/19/2009.

Our data collection period spanned the most recent 2021-22 and 2022-23 seasons. We used the following dates, starting at 00:00 EST and ending at 23:59 EST, to represent the regular season and the NCAA tournament:

- 2021-22 regular season
  - Men's: 11/9/2021 – 3/13/2022
  - Women's: 11/9/2021 – 3/6/2022
- 2021-22 NCAA tournament
  - Men's: 3/15/2022 – 4/4/2022
  - Women's: 3/18/2022 – 4/3/2022
- 2022-23 regular season
  - Men's and women's: 11/7/2022 – 3/12/2023
- 2022-23 NCAA tournament
  - Men's: 3/14/2023 – 4/3/2023
  - Women's: 3/15/2023 – 4/2/2023

Using the Twitter API /2/tweets/search/all endpoint (also known as full archive search) and user IDs, we fetched tweets for each season that were created between the start of the regular season and the end of the NCAA tournament. In total, we downloaded 1.1 million tweets via the Twitter API from 5/4/2023-5/6-2023. Pulling tweets in May after the 2023 season ended presented a few limitations. First, teams might have changed their Twitter accounts between the two seasons. Second, the 2021-22 season tweets had a longer opportunity to acquire likes, retweets, quotes, and replies before we downloaded the public metrics. The first limitation affected 6 out of 724 teams. Addressing the last limitation, recent research has shown that no relevant number of impressions can be observed for ∼95% of tweets after 24 hours [35], suggesting the two seasons' numbers can be compared.

Following prior work [36], we parameterized our full archive search requests to include several additional tweet fields, including the author ID, when the tweet was created, public metrics (e.g., like counts, retweet counts, etc.), referenced tweet data if applicable (e.g., if this tweet is a retweet then the author ID of the original tweet, if this tweet is a reply then the author ID of the tweet replied to, etc.), and tagged entity data (e.g., mentions, hashtags, etc.). With this extra tweet information, we were able to perform interesting

[2]https://developer.twitter.com/
[3]https://web3.ncaa.org/directory/

analyses of the tweets, such as using the tweet creation time to assign which part of the season the tweet was published in.

Once we fetched the team tweet data, we turned our attention to collecting each season's schedule and game outcome information for each team. To do this, we scraped public facing ESPN webpages that contained a list of men's college basketball teams[4] and women's college basketball teams,[5] grouped by conference. Upon page inspection, each team's unique ESPN ID was obtained from the team's ESPN page link. For example, clicking on the men's Gonzaga Bulldogs link redirects to their ESPN team page: `https://www.espn.com/mens-college-basketball/team/_/id/2250/gonzaga-bulldogs`. From this URL, we extracted the ESPN ID for the Gonzaga Bulldogs (2250). This ID represents the institution and is therefore the same for both men's and women's teams at the same institution. With each team's ESPN ID, we were able to visit and scrape schedule information for each season of interest. For example, the men's Gonzaga Bulldogs 2022-23 schedule and game outcome information were available on the webpage accessible by replacing the display name `gonzaga-bulldogs with season/2023` in the above URL.

Following the game detail links on the schedule pages, we extracted attributes about each game, including the game time, opponent, outcome (e.g., win, lose, tie, forfeit, etc.), score, and team ranks. After we processed all team schedules, we aligned the Twitter user IDs with the ESPN IDs. This was performed manually by comparing Twitter handles/descriptions with ESPN display names. As an example of the alignment process, the Gonzaga University men's basketball team's Twitter handle (at the time of writing) was @ZagMBB, while their Twitter ID was 602989093, ESPN display name was Gonzaga Bulldogs, and their ESPN ID was 2250. Once aligned, we could merge the Twitter data and the ESPN data. To do this, each tweet was aligned by team ID, then timestamp aligned by tweet creation and game times from ESPN. To prevent misalignment due to different time zones, all timestamp alignment was performed using UTC localized timestamps.

After collecting, cleaning, and aligning data from Twitter and ESPN, we performed exploratory data analysis to determine the structure and content of the team tweets and social networks at a high level. This included calculating simple summary statistics such as counts and means for tweet public metrics (e.g., likes, retweets, quote tweets, and replies) and tweet references (e.g., mentions, hashtags, and links). For further exploration, we grouped the tweets by time (e.g., part of season: regular, tournament, or in between; day of week), team characteristics (e.g., gender and tournament performance), and conference (e.g., SEC, Big Ten, etc.). While the dataset was rich enough to support additional interesting groupings, such as by regular season rankings, in this paper we present the analyses most relevant to social network modeling and tweet-based game winner prediction.

Next, we modeled and visualized team interactions on Twitter via social networks. Each social network is comprised of nodes (in our case, teams) and edges that connect the nodes (in our case, interaction on Twitter). We calculated interaction following prior work that created a social influence network using official Twitter accounts from U.S. Congressional members [37]. This approach creates a weighted, bi-directional network by measuring "influence" between two Twitter accounts. Influence is the interaction between the two teams via retweeting, quoting, replying, and mentioning. When constructing the influence network, we only included teams if they created at least 100 tweets during the time period of interest. This inclusion criterion helped ensure the empirically calculated edge weights were meaningful. For the 2022-23 season, there were 9 men's teams and 13 women's teams that did not meet this inclusion criterion.

B. GAME WINNER PREDICTION

Besides tweet metrics, team rankings, and game scores, the data we collected in the days leading up to a college basketball matchup were primarily text-based in the form of tweets. To predict the winner of a college basketball matchup using these tweets, we aligned the two team's recent tweets at the game level. For each game, we identified the game time and which two teams were playing. For each team, we then identified their N most recent tweets that were created before the game started. During this process, we performed minimal pre-processing of the tweets, including replacing links with "<URL>" and appending the tweets together using spaces. Each team's N-tweets were then paired (the order of which team's N-tweets were presented first was chosen randomly) and matched with the game outcome. The recent N-tweets pairs could then be used as input to LLMs. To leverage the established high performance of XGBoost for March Madness prediction, we utilized LLMs with two different approaches to next game winner prediction:

1) Embeddings + XGBoost: using an LLM to get numeric embeddings for each team's N-tweets. The embeddings were then combined and used as input into a XGBoost classification model trained to predict game winners.
2) In-context learning [38]: using the N-tweets pairs as context when constructing a prompt that asks an LLM to predict a game winner. We did this using:
    a. Zero-shot learning (no examples in the prompt)
    b. Few-shot learning (K examples in the prompt)

For both approaches, we utilized LLMs from OpenAI using their API.[6] For approach #1, we utilized OpenAI's `text-embedding-ada-002` model (8,191 token limit; 1,536 output dimensions) to produce embeddings for each team's N-tweets in a game matchup. Embeddings are numeric representations of text such that two texts that are related will have numeric representations that are close to each other in a high

[4] https://www.espn.com/mens-college-basketball/teams

[5] https://www.espn.com/womens-college-basketball/teams

[6] https://platform.openai.com/

```
### Instruction: Two college basketball teams are about to play a game.
Given each team's recent tweets, predict which team will win.

### Examples:
Example: Coastal Carolina Chanticleers Men vs Appalachian State
Mountaineers Men
Coastal Carolina Chanticleers Men tweets: You already know what time it
is⌛ <URL> Here are our starters for tonight's action!! □: Conway, S.C. □:
HTC Center ⏰ 7:00 p.m. ET □ <URL> □: <URL> □: <URL>
#TEALNATION | #CHANTSUP <URL> We're 60 minutes away from the
opening tip...You have time to make it over to the HTC Center! □: Conway,
S.C. □: HTC Center ⏰ 7:00 p.m. ET □ <URL> □: <URL> □: <URL>
#TEALNATION | #CHANTSUP; #TealNation we need you guys to come
support during our home stand. Tonight 7:00 pm □ Appalachian State
Saturday 2:00 pm □ South Alabama Monday 7:00 pm □ Chicago State
<URL> Hey Chanticleer fans...We need you at the HTC tonight when
Appalachian State comes to town...CCU There! □: Conway, S.C. □: HTC
Center ⏰ 7:00 p.m. ET □ <URL> □: <URL> □: <URL> #TEALNATION |
#CHANTSUP <URL> Men's Basketball Back Home to Host a Three-Game
Homestand <URL> Happy Birthday to Freshman Guard
@marcus_saunders_ ‼□□ <URL> Men's Basketball Falls on the Road at
Georgia State on Saturday Afternoon #TEALNATION | #CHANTSUP
<URL>
Appalachian State Mountaineers Men tweets: Tonight's starting □ Tune in
on ESPN+: <URL> #TakeTheStairs <URL> It's GAME DAY! We are set to
visit Coastal Carolina at 7 p.m. on ESPN+. □ <URL> (ESPN+) □ <URL>
□ <URL> #TakeTheStairs <URL> RT @appstatesports: Wins across all □
competing sports last weekend. Let's do it again. #GoApp <URL>
Previewing tomorrow night's contest at Coastal Carolina □ #TakeTheStairs
<URL> Mic'd up □□ with Assistant Coach Patrick Moynihan
#TakeTheStairs <URL> Please join us in wishing a happy birthday to
Xavion Brown! #TakeTheStairs <URL> Xavion Brown will be live on
Mountaineer Talk tonight at 7:30 along with Coach Kerns! #TakeTheStairs
<URL> <URL> RT @appstate: Join App State in honoring the life and
legacy of Dr. King in 2023. Learn more □ <URL> <URL>
Winner: Coastal Carolina Chanticleers Men

<2 examples removed for brevity>

### Question:
Now predict which team will win this game: Texas A&M-Corpus Christi
Islanders Men vs Alabama Crimson Tide Men
Texas A&M-Corpus Christi Islanders Men tweets: LUTZ GO ISLANDERS !!!!
□ No. 1 Alabama □ Legacy Center (Birmingham, Ala.) ⏰2:15 p.m. □ CBS
□ <URL> □ <URL> □ <URL> #ShakasUp // #LutzGo // #TAMUCC //
#MarchMadness <URL> @WarDamnStool □ <URL> Lutz Shakas the
World □ No. 1 Alabama □ Legacy Center (Birmingham, Ala.) ⏰1:45 p.m.
□ CBS □ <URL> □ <URL> □ <URL> #ShakasUp // #LutzGo // #TAMUCC
// #MarchMadness <URL> RT @Go_Islanders: Get up Islanders, It's Game
Day! □ #LutzGo #ShakasUp□ #MarchMadness □ <URL> RT
@CBKReport: Texas A&amp;M Corpus Christi's campus is ridiculous
<URL> RT @CallerSports: Still Dancing: @Islanders_MBB makes
history with NCAA Tournament win, prepares for No. 1 Alabama.
#MarchMadness https:/…; RT @qmartinez: De'Lazarus Keys and Simeon
Fryer rode out the struggles of their first season with @Islanders_MBB
stuck with the program,…; RT @LarissaLiska: I am #MarchMadness ready!
□ #FirstRound matchup between @Islanders_MBB and Alabama is
Thursday at 1:45 p.m. Catch the wat…
Alabama Crimson Tide Men tweets: Starting □ against Texas A&amp;M
CC. #RollTide | #BlueCollarBasketball <URL> □ROLL CALL Alabama vs
Texas A&amp;M CC in the 1st Round of the NCAA Tournament! Where
are you watching the game from?! #RTR <URL> Here comes the Tide!
#RollTide | #BlueCollarBasketball <URL> RT @MarchMadnessMBB: □
"ROLL TIDE" Chuck is going with @AlabamaMBB as his pick for National
Champion □ @NBAonTNT #MarchMadness https://t…; Let's Dance □
<URL> RT @NaismithTrophy: The only player to win @SEC Player of the
Year, Freshman of the Year and Tournament MVP in the same season □
Brandon…; A new chapter begins today. #RollTide | #BlueCollarBasketball
<URL> Game 35: The Tide &amp; Texas A&amp;M CC face off in
Birmingham! #RollTide | #BlueCollarBasketball <URL>

Use format Winner: ...
```

FIGURE 2. An example few-shot learning prompt for the 2023 tournament with N = 8 and K = 3 (1 example shown).

dimensional space. We combined the two N-tweets embeddings to produce a feature vector that could serve as input to a machine learning classifier. Specifically, we used the Python implementation of XGBoost trained with logistic regression for binary classification as the objective and 50 early stopping rounds on all 2022 games and regular season, non-tournament 2023 games. One third of the training games were held out for use as the validation set. Using this set, we explored various ways to combine the two teams' embeddings, such as concatenation, averaging, and subtraction, experimentally finding concatenation yielded the best results. The model was evaluated with the 134 men's and women's tournament games from 2023. We included the First Four play-in games to increase the number of test instances used for evaluation.

For approach #2, prompt engineering can be applied to create prompts from N-tweets pairs for in-context learning with LLMs in several ways. First, when paired with an instruction such as "Two college basketball teams are about to play a game. Given each team's recent tweets, predict which team will win.", the recent N-tweets pairs can be used directly as prompts with zero-shot learning. For few-shot learning, the instruction can be modified to inform the LLM that examples will be included. Then, K recent N-tweets pairs/game outcomes can be provided as examples of the task. For clarity, Figure 2 shows an example few-shot prompt with N = 8 and K = 3 (2,406 total tokens). Since our prediction task focused on the 2023 tournament games, examples were chosen randomly from the 2022-23 season regular games that had N recent tweets for both teams. For many teams, this inclusion criterion eliminated their first few games.

To elaborate, N and K are parameters that must be adjusted depending on the LLM used. The maximum input length of an LLM is measured in terms of tokens. OpenAI LLMs have a long enough maximum input token length for our proposed approach with few-shot learning. It is worth noting that the N-tweets pairs can be used to further train LLMs as part of a fine-tuning task. In this work, we primarily explored in-context learning because LLMs are already familiar with the core concepts of our prediction task (e.g., basketball, tweets, etc.) and we wanted to evaluate the viability of using off-the-shelf LLMs without explicit training, which increases the likelihood that this approach could be adopted without a significant investment in LLM training resources. For in-context learning, we utilized OpenAI's `gpt-3.5-turbo` and `gpt-3.5-turbo-16k models` (4,097 and 16,385 token limits, respectively) with temperature set to 0.0 and the following system message to help set the behavior of the model: "You are an expert in sports analytics. Please follow instructions to forecast college basketball game winners." According to OpenAI, these models were trained on data up to September of 2021[7] and therefore have not been exposed to the 2021-22 or 2022-23 basketball season game results or team tweets; however, these models have likely distilled some historical knowledge about the teams/institutions, and this should be kept in mind when interpreting their outputs.

Using the OpenAI API, we requested predicted completions for all 134 men's and women's NCAA tournament games in the 2022-23 season. To determine the winning team name from the predicted completion, a text similarity

[7]https://platform.openai.com/docs/models/gpt-3-5

TABLE 1. **Men's tweet public metrics and tweet content references.**

	2021-22	*2022-23*	*% Change*
# Tweets	291,996	291,277	-0.25%
Avg # Likes	83.48	72.78	-12.82%
Avg # Retweets	6.95	13.07	-22.89%
Avg # Quotes	1.69	1.46	-13.26%
Avg # Replies	1.77	1.93	8.96%
Avg # Mentions in Tweet	0.64	0.62	-2.39%
Avg # Hashtags in Tweet	0.89	0.86	-3.98%
Avg # Links in Tweet	0.98	1.03	5.57%

TABLE 2. **Women's tweet public metrics and tweet content references.**

	2021-22	*2022-23*	*% Change*
# Tweets	248,730	268,729	8.04%
Avg # Likes	36.25	39.39	8.68%
Avg # Retweets	7.33	7.77	6.06%
Avg # Quotes	0.78	0.74	-4.88%
Avg # Replies	0.55	0.65	17.89%
Avg # Mentions in Tweet	0.53	0.51	-3.28%
Avg # Hashtags in Tweet	0.96	0.95	-0.95%
Avg # Links in Tweet	1.03	1.07	4.41%

measure based on the Levenshtein Edit distance was computed for each team in the matchup and the predicted completion. Of the two teams, the team with the highest similarity result was considered the predicted winner. We evaluated the prediction approach using the simple accuracy metric (# of correct predictions / # of total predictions) with comparisons to a few baselines. First, from a statistical perspective, predicting the winner of a game is a binary classification task where each team is equally likely to win if no prior knowledge is used in the prediction. Using the inverse survival function of a binomial distribution, we can determine a critical value, p_{critical}, at which n Bernoulli trials are expected to exceed 0.5 at a significance level α [39]. If a model's prediction accuracy $\geq p_{\text{critical}}$, then we can conclude the model performed significantly better than random and the two teams' recent tweets hold some predictive power. For 134 games and $\alpha = 0.05$, $p_{\text{critical}} = 0.575$. Second, we considered a baseline prediction model that always chooses the team with the higher seed as the winner of each game. Next, to account for models' prior learned notions of the teams from training, we requested `gpt-3.5-turbo` predictions for zero-shot/zero-context learning as well, meaning the prompts contained only the team names (no examples from other teams and no team tweets as context). The final baseline predictions we used were from the 2023 Kaggle Competition winning submission [8]. From the winning submission's predicted probabilities, we determined the predicted winning team using a 0.5 probability threshold.

Finally, we leveraged the ability of LLMs to explain their reasoning process to provide insight about what hidden information and/or social network interaction in an official team tweet is responsible for its predictive power. Chain of thought prompting [40] is a promising prompt engineering approach to achieve this because it has been shown to increase the explainability of domain-specific predictions [21]. To apply chain of thought prompting, we modified the end of each prompt to include, "Let's think step by step.\nUse format Reasoning process: ...\nWinner: ..." (see Figure 2 for an example of the original prompt).

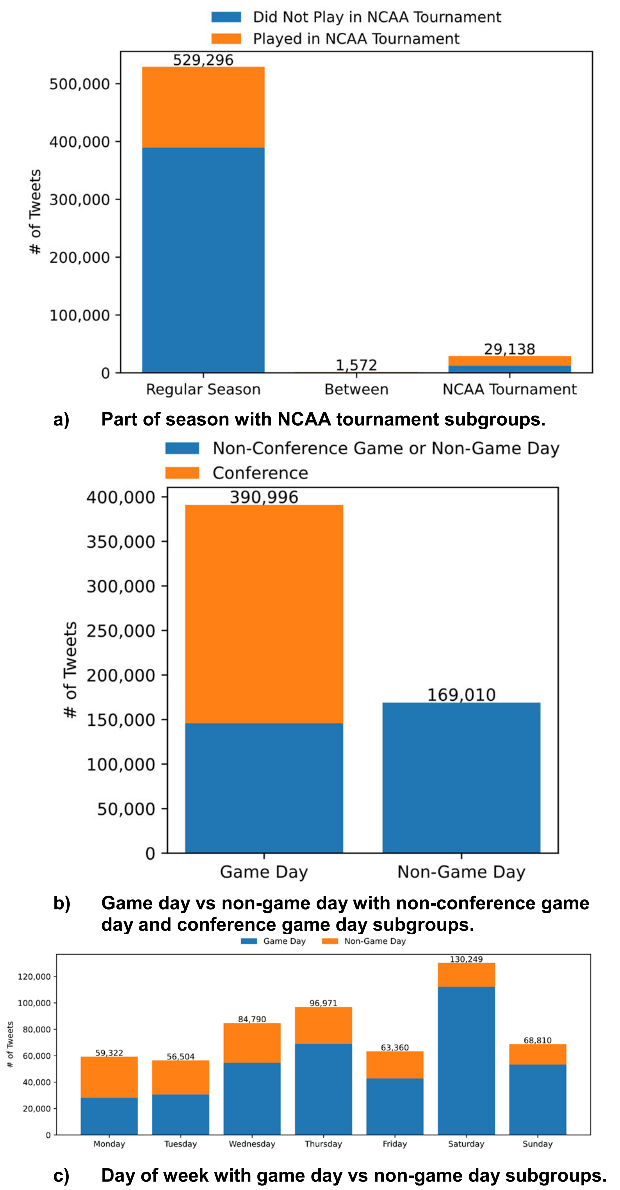
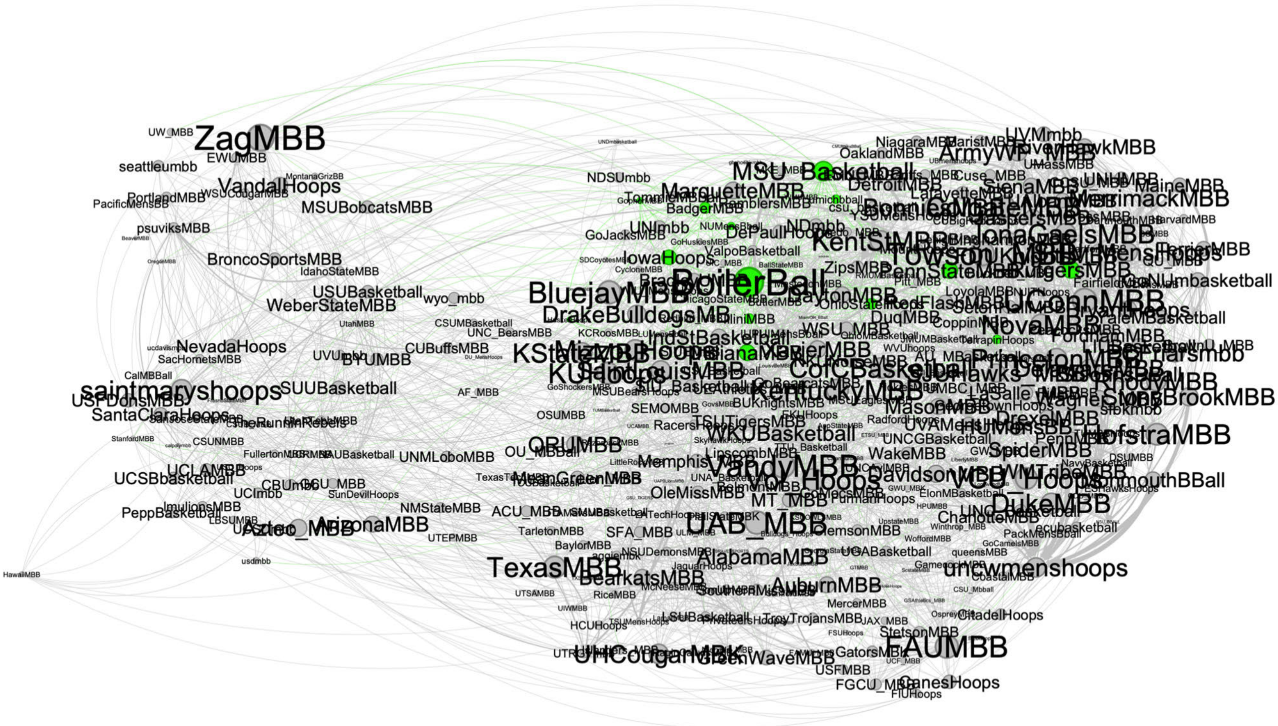


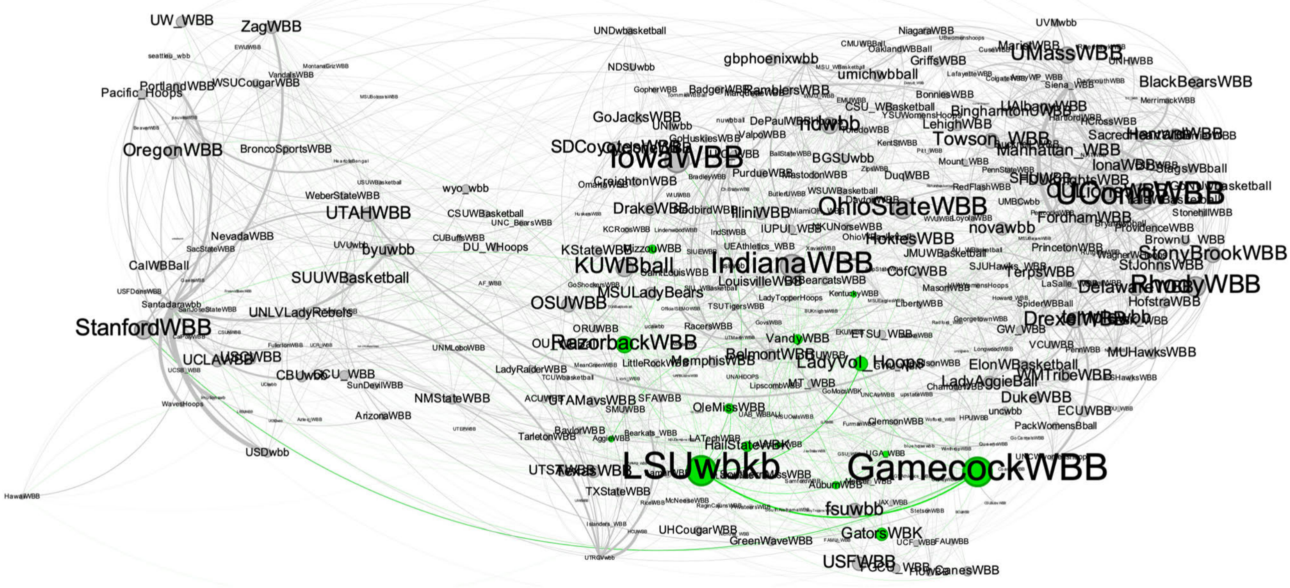
FIGURE 3. **The number of tweets by part of the season and by game day for the 2022-23 season.**

IV. RESULTS

Beginning with summary statistics, 540,726 tweets were created during the 2021-22 season and 560,006 tweets were created during the 2022-23, representing a 3.57% increase. We break these seasonal differences down by team gender,



a) Men's social network with the Big Ten conference highlighted in green.



b) Women's social network with the SEC conference highlighted in green.

FIGURE 4. Social media influence networks for the 2022-23 season. Team nodes are positioned using GPS coordinates with a no-overlap policy to spread the nodes out. Team node size is relative to out degree and edge thickness is relative to influence.

with men's in Table 1 and women's in Table 2. These tables also provide per tweet averages for public metrics and tweet content references. When interpreting seasonal comparisons, it is worth noting again that there were fewer teams in our 2021-22 season subset due to division reclassification and Twitter account changes. Next, Figure 3.a shows the number of tweets in three different parts of the season, Figure 3.b shows the number of tweets on game days versus non-game

TABLE 3. Game outcome prediction accuracy for the 2023 NCAA tournament games.

		Overall	*Men's*	*Women's*
#1: Embed + XGBoost	***N = 8***	61.9%	**65.7%**	58.2%
	N = 16	59.7%	62.7%	56.7%
#2a: Zero-shot	***N = 8***	63.4%	56.7%	70.1%
	N = 16	63.4%	59.7%	67.2%
#2b: Few-shot	***N = 8***	**64.9%**	59.7%	**70.2%**
	N = 16	63.4%	58.2%	68.7%
Baselines	***ZS/ZC***	54.1%	58.6%	49.1%
	HS	71.4%	*69.8%*	73.0%
	Kaggle	*73.1%*	65.7%	*80.6%*

N = # of few-shot examples, ZS/ZC = zero-shot/context, HS = higher seed, **bold** indicates best LLM accuracy, *italics* indicates best accuracy overall.

Reasoning process:
To predict the winner of the game between Texas A&M-Corpus Christi Islanders Men and Alabama Crimson Tide Men, we will analyze the recent tweets from both teams.

Texas A&M-Corpus Christi Islanders Men tweets mention their game against No. 1 Alabama, the location, time, and broadcasting details. They also show excitement and support from their fans. However, there is no specific information about their recent performance or any notable achievements.

Alabama Crimson Tide Men tweets mention their game against Texas A&M CC in the 1st Round of the NCAA Tournament. They use hashtags related to their team and basketball culture. They also have a tweet from a prominent basketball personality picking them as the National Champion. However, there is no specific information about their recent performance or any notable achievements.

Based on the available information, it is difficult to determine a clear winner. However, considering that Alabama is ranked No. 1 and has received support from a basketball personality, they might have a stronger team.

Winner: Alabama Crimson Tide Men

a) Response for the example provided in Figure 2.

Reasoning process:
1. Analyze the recent tweets of each team to gather information about their mindset, preparation, and confidence.
2. Look for any specific mentions of the game, such as game time, location, or opponent.
3. Consider any notable achievements or honors mentioned in the tweets.
4. Assess the overall tone and energy of the tweets to gauge the team's enthusiasm and motivation.

Winner: Based on the analysis of the tweets, it appears that both teams are excited and focused on the game. However, the Iowa Hawkeyes Women's tweets mention the "WFinalFour" multiple times, indicating that they are competing in a significant tournament. This suggests that the Iowa Hawkeyes Women may have a higher level of motivation and determination. Therefore, the predicted winner of the game is Iowa Hawkeyes Women.

b) Response for the 2022-23 women's championship game. Note that the prediction is incorrect.

FIGURE 5. Example chain of thought reasoning responses.

days, and Figure 3.c shows the number of tweets per day of the week.

To model and visualize the interaction among the teams on Twitter, we generated two social influence networks, one for the men's teams and one for the women's teams. There were 354 nodes and 2,702 edges in the men's network and the top three influencers in the network by node out-degree were @BoilerBall (23), @ZagMBB (23), and @VandyMBB (21). Figure 4.a shows a graph visualization of the men's network with @BoilerBall and its conference (Big Ten) highlighted. On the women's side, there were 344 nodes and 2,296 edges and the top three influencers in the network by node out-degree were @GamecockWBB (32), @LSUwbkb (out-degree 32), and @UConnWBB (out-degree 27). Figure 4.b shows a graph visualization of the women's network with @LSUwbkb and its conference (SEC) highlighted.

For game winner prediction, we explored values of N and K that would fit within the `gpt-3.5-turbo` max token length. We found N = 8 and K = 3 were the largest values that produced all 134 game prompts under the `gpt-3.5-turbo` token limit. For context, the average men's team produced 5.46 tweets per day and the average women's team produced 5.11 tweets per day. This suggests N = 8 represents about a day and a half of recent team tweets, though as Figures 3.b and 3.c show, tweeting behavior is not consistent from day to day. With `gpt-3.5-turbo-16k`, we could provide even more context to the model for prediction. Consequently, we ran a second experiment with double the number of recent tweets (N = 16), representing about three days. For comparison, we also used N = {8, 16} for our embeddings + XGBoost (approach #1) experiments, yielding 20,371 (N = 8) and 20,047 (N = 16) games for training/validation.

Table 3 provides the accuracies from these experiments and from our baseline prediction approaches. We do not report accuracies for the chain of thought prompting experiments because the responses occasionally contained predicted winner text such as "Cannot determine" or "Based on the analysis of the recent tweets, it is difficult to determine a clear winner." We do however provide two example chain of thought prompt responses containing valid predictions in Figure 5. Finally, it is worth noting that for the zero-shot/zero-context learning experiment, the model would sometimes return a safe-guarded non-answer, such as "…It is not possible to accurately predict the winner of a specific college basketball game without considering various factors such as …Additionally, as an AI language model, I do not have access to real-time data or the ability to analyze current team dynamics. Therefore, I cannot provide a definitive answer…" In these cases, we forced a predicted winner by choosing a winning team at random.

V. DISCUSSION

In this paper, we investigated the predictive power of men's and women's Division I college basketball social network data. Specifically, we aimed to use tweets and LLMs from OpenAI to predict which team will win in a game matchup during the 2022-23 NCAA tournament.

A. TWEET ACTIVITY AND SOCIAL NETWORK INFLUENCE

Beginning with the results in Tables 1 and 2, it appears that the official team Twitter account usage between seasons exhibited a different trend for the men's teams versus the women's teams. The total number of tweets for the men's teams was surprisingly stable from the 2021-22 season to 2022-23 season; however, their reception by the Twitter community appeared to become less impactful, with negative percent changes for like, retweet, and quote tweet averages. The women's public metric stats told a conflicting story, with an increase in number of tweets and average likes and retweets per tweet. While we cannot infer the cause of these differences directly, we suspect the increased spectatorship of women's sports during the NCAA tournament due to the record-breaking performance of Iowa's Caitlyn Clark impacted the numbers in some fashion. As for the structured content of the tweets, both teams appear to be reducing their use of mentions and hashtags, while increasing their use of links. The links could be referencing a variety of media types, such as websites, images, GIFs, videos, etc. Following these links and replacing the <URL> placeholder with link content information represents a potential direction for future work.

Figure 3.a shows when the teams were tweeting (regular season, between the regular season and tournament, and during the tournament). While the total tweet numbers were not surprising given the duration of these timeframes, it was interesting that the teams that made it to the tournament appeared to tweet more than their fair share during the regular season. Figures 3.b and 3.c show when these tweets were happening relative to when the teams were playing. Clearly teams tweeted more on game days (and particularly on Saturdays) while still being quite active on non-game days. Lastly, Figures 4.a and 4.b show there was a lot of interaction amongst the men's and women's teams on Twitter during the 2022-23 season. In each network, the highlighted nodes/edges represent the teams/interactions within the conference with the most influential team in the network. Following the edges from these highlighted nodes shows there was a high amount of interaction amongst teams within and outside of their conference, particularly in the denser men's network. The tweets used for game winner prediction capture these interactions as mentions, replies, retweets, and quote tweets in the days leading up a game matchup between two teams, offering insight that social media activity can be useful for game winner prediction.

B. GAME WINNER PREDICTION

For winner prediction of the 134 tournament games in 2023, we considered a prediction accuracy greater than 57.5% to be better than random (based on p_{critical}). All but two of the experimental results in Table 3 using LLMs exhibited accuracies greater than this cutoff, indicating both LLM approaches, approach #1 (embeddings + XGboost) and approach #2 (in-context learning), hold predictive power for this social media-based forecasting task. As expected, the winning Kaggle Competition submissions for the 2023 tournament produced the best accuracy with 73.1% overall, performing particularly well for the women's tournament (80.6%). When comparing the men's accuracies to the women's accuracies, except for approach #1 most of the methods performed better for the women. This is not surprising since the women's tournament tends to experience fewer upsets. Note that the Kaggle accuracies are not directly comparable to our tweet-based accuracies for a few reasons. First, the Kaggle predictions were made in advance of the tournament and therefore did not have access to timely information on social media. Second, the Kaggle Competition was scored by the Brier score, not accuracy like we are reporting.

The best overall performing LLM approach for men's and women's teams was few-shot learning with N = 8 (accuracy 64.9%), although for the men's games, both embeddings + XGBoost experiments outperformed in-context learning, with N = 8 matching the Kaggle accuracy. This suggests that both LLM approaches demonstrate predictive merit, especially when used with fewer recent N-tweets. Between the two similarly performing approaches, one reason to favor few-shot learning over embeddings + XGBoost is it uses tweet text directly and does not require a feature engineering step.

Zooming in on approach #2, few-shot learning generally outperformed zero-shot learning and it outperformed zero-shot/zero-context by about 10%, proving that tweet context was predictive of next game performance; however, few-shot learning was barely better than zero-shot/zero-context for the men's teams (~1%) and quite a bit better for the women's teams (~20%). The poorer zero-shot/zero-context accuracy for women's teams suggest that the original LLM training set likely contained more historical performance information about the men's teams than the women's teams.

Though we were limited by model token lengths, we explored two values of N to see the effects of historical tweet context length on prediction accuracy. For in-context learning, it appears that a shorter tweet context (N = 8) led to a slightly higher accuracy than a longer tweet context (N = 16). Future work aims to explore additional values of N, as well as K (the number shots used with in-context learning). The latter is important because the choice of shot number can be tuned to achieve greater LLM performance [41], with 4 – 8 exemplars generally yielding good results [42].

Chain of thought prompting provided insights into the reasoning processes LLMs employed when making predictions. The enumerated reasons provided in the response shown in Figure 5.b suggest showing confidence, mentioning/interacting with the opponent, and conveying an energetic tone are signs a team is likely to win their upcoming game. The response in Figure 5.a is less structured and suggests that tweets containing spirited hashtags, notable achievements, and support from credible sources such as a "prominent basketball personality" are informative. These diverse predictive qualities are all embedded in social network data, eliminating the need to collect this information from disparate sources.

C. LIMITATIONS AND FUTURE WORK

While this work provides promising prediction results, there are several limitations to discuss. First, this technique cannot be used to make March Madness brackets before the first tournament game. This is because, by design, the recent N-tweets pairs used for inference are not available until shortly before a game starts. On the other hand, these just-in-time predictions can be useful for sports betting or posting predictions on social media, without the need for deploying a custom trained model (e.g., an inference API like OpenAI or a chat interface like ChatGPT can be used to produce game winner predictions directly from social media text). Another limitation includes the recent transition from Twitter to X. With this change, Twitter removed their generous API access and replaced it with several pay-walled access tiers, reducing the likelihood of future social media research (of any kind) with their platform. While this work as presented used Twitter data and its API, it is a general game outcome prediction framework that can use text from team posts on any social media platform.

Given the novelty of this work, there are several promising directions for future exploration. First, recent prompt engineering techniques such as iterative bootstrapping in chain of thought prompting [43], self-consistency [44], generated knowledge prompting [45], [46], or least-to-most prompting [47] can be leveraged to improve the few-shot learning results. While we briefly explored chain of thought prompting in this paper, a more thorough analysis of what social media information LLMs find predictive would shed more light on the reasons why LLMs predict certain teams to win. Additionally, once this information is identified, it could be extracted and fused with the more traditional team/player performance and expert opinion features to increase the predictive power of state-of-the-art models. Given the success of a similar approach by Beal and colleagues, this performance gain seems likely [34]. Finally, though we did not pursue fine tuning LLMs as a third prediction approach, perhaps parameter efficient fine tuning with our in-context learning prompts [38] may offer an improvement in accuracy as well.

VI. CONCLUSION

In this paper, we presented a large-scale social media dataset containing the tweets of all men's and women's Division I college basketball teams during the most recent two seasons. We used this dataset to explore the first application of social network data and in-context (few-shot/zero-shot) learning for sporting event outcome prediction. While the goal of this work was not to beat the state-of-the-art prediction models with tweets alone, our work has provided evidence of the predictive power of official team tweets, interactions amongst teams in a social network, and LLMs for forecasting game outcomes. For the 2022-23 season, embeddings + XGBoost yielded a 65.7% prediction accuracy for the men's tournament games and few-shot learning yielded a 70.2% prediction accuracy for the women's tournament games. These accuracies support future work exploring the application of LLMs in the field of sports analytics.

ACKNOWLEDGMENT

The authors wishes to thank Andrew Kim and Jimmy Smith for college basketball insights, as well as Chris Fink for social network advice and Diane Cook for LLM prompting ideas.

REFERENCES

[1] F. J. R. Ruiz and F. Perez-Cruz, "A generative model for predicting outcomes in college basketball," *J. Quant. Anal. Sports*, vol. 11, no. 1, pp. 39–52, Jan. 2015.

[2] J. Gumm, A. Barrett, and G. Hu, "A machine learning strategy for predicting March madness winners," in *Proc. IEEE/ACIS 16th Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput. (SNPD)*, Jun. 2015, pp. 1–6.

[3] A. Hoegh, M. Carzolio, I. Crandell, X. Hu, L. Roberts, Y. Song, and S. C. Leman, "Nearest-neighbor matchup effects: Accounting for team matchups for predicting March Madness," *J. Quant. Anal. Sports*, vol. 11, no. 1, pp. 29–37, Mar. 2015.

[4] M. J. Lopez and G. J. Matthews, "Building an NCAA men's basketball predictive model and quantifying its success," *J. Quant. Anal. Sports*, vol. 11, no. 1, pp. 5–12, Mar. 2015.

[5] S. Dutta, S. H. Jacobson, and J. J. Sauppe, "Identifying NCAA tournament upsets using balance optimization subset selection," *J. Quant. Anal. Sports*, vol. 13, no. 2, pp. 79–93, Jun. 2017.

[6] I. G. Ludden, A. Khatibi, D. M. King, and S. H. Jacobson, "Models for generating NCAA men's basketball tournament bracket pools," *J. Quant. Anal. Sports*, vol. 16, no. 1, pp. 1–15, Mar. 2020.

[7] J. W. Kim, M. Magnusen, and S. Jeong, "March Madness prediction: Different machine learning approaches with non-box score statistics," *Managerial Decis. Econ.*, vol. 44, no. 4, pp. 2223–2236, Jun. 2023.

[8] J. Sonas and W. Cukierski. *March Machine Learning Mania 2023*. Kaggle. Accessed: Oct. 30, 2023. [Online]. Available: https://kaggle.com/competitions/march-machine-learning-mania-2023

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2016, pp. 785–794.

[10] *The Longest an NCAA Bracket Has Ever Stayed Perfect | NCAA.com*. Accessed: Oct. 31, 2023. [Online]. Available: https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-17/longest-ncaa-bracket-has-ever-stayed-perfect

[11] *Why the @Raddar Solution Works*. Accessed: Oct. 30, 2023. [Online]. Available: https://kaggle.com/code/nigelhenry/why-the-raddar-solution-works

[12] *The Absurd Odds of a Perfect NCAA Bracket | NCAA.com*. Accessed: Oct. 30, 2023. [Online]. Available: https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-16/perfect-ncaa-bracket-absurd-odds-march-madness-dream

[13] *NCAA Tournament Bracket Challenge 2023*. ESPN. Accessed: Oct. 30, 2023. [Online]. Available: https://fantasy.espn.com/tournament-challenge-bracket/2023/en/entry?entryID=89768899

[14] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, "Emergent abilities of large language models," 2022, *arXiv:2206.07682*.

[15] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, L. Zhao, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, and B. Ge, "Summary of ChatGPT-related research and perspective towards the future of large language models," *Meta-Radiology*, vol. 1, no. 2, Sep. 2023, Art. no. 100017.

[16] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.

[17] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, and G. Kasneci, "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individual Differences*, vol. 103, Apr. 2023, Art. no. 102274.

[18] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature Med.*, vol. 29, no. 8, pp. 1930–1940, Aug. 2023.

[19] H. Xue and F. D. Salim, "PromptCast: A new prompt-based learning paradigm for time series forecasting," 2022, *arXiv:2210.08964*.

[20] H. Xue, F. D. Salim, Y. Ren, and C. L. A. Clarke, "Translating human mobility forecasting through natural language generation," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 1224–1233.

[21] X. Yu, Z. Chen, Y. Ling, S. Dong, Z. Liu, and Y. Lu, "Temporal data meets LLM—Explainable financial time series forecasting," 2023, *arXiv:2306.11025*.

[22] M. Latififard, M. Saffari, R. Norouzi Seyed Hossini, and M. Bahrani, "A systematic review of sports-related studies on social media," *Commun. Res.*, vol. 27, no. 102, pp. 151–169, Aug. 2020.
[23] B. R. Haugh and B. Watkins, "Tag me, tweet me if you want to reach me: An investigation into how sports fans use social media," *Int. J. Sport Commun.*, vol. 9, no. 3, pp. 278–293, Sep. 2016.
[24] B. Browning and J. Sanderson, "The positives and negatives of Twitter: Exploring how student-athletes use Twitter and respond to critical tweets," *Int. J. Sport Commun.*, vol. 5, no. 4, pp. 503–521, Dec. 2012.
[25] J. L. David, M. D. Powless, J. E. Hyman, D. M. Purnell, J. A. Steinfeldt, and S. Fisher, "College Student athletes and social media: The psychological impacts of Twitter use," *Int. J. Sport Commun.*, vol. 11, no. 2, pp. 163–186, Jun. 2018.
[26] R. P. Schumaker, A. T. Jarmoszko, and C. S. Labedz, "Predicting wins and spread in the premier league using a sentiment analysis of Twitter," *Decis. Support Syst.*, vol. 88, pp. 76–84, Aug. 2016.
[27] H. Wäsche, G. Dickson, A. Woll, and U. Brandes, "Social network analysis in sport research: An emerging paradigm," *Eur. J. Sport Soc.*, vol. 14, no. 2, pp. 138–165, Apr. 2017.
[28] K. E. Pearson. *Understanding How the Southeastern Conference Football Teams Use Twitter Through a Content Analysis*. Accessed: Oct. 30, 2023. [Online]. Available: https://scholarworks.uark.edu/etd/1581
[29] S. S.-C. Chen, T. Duncan, E. Street, and B. Hesterberg, "Differences in official athletic website coverage and social media use between men's and women's basketball teams," *Sport J.*, vol. 49, Jul. 2016, Accessed: Oct. 30, 2023. [Online]. Available: https://thesportjournal.org/article/differences-in-official-athletic-website-coverage-and-social-media-use-between-mens-and-womens-basketball-teams/
[30] L. Wallace, J. Wilson, and K. Miloch, "Sporting facebook: A content analysis of NCAA organizational sport pages and big 12 conference athletic department pages," *Int. J. Sport Commun.*, vol. 4, no. 4, pp. 422–444, Dec. 2011.
[31] N. UzZaman, R. Blanco, and M. Matthews, "TwitterPaul: Extracting and aggregating Twitter predictions," 2012, *arXiv:1211.6496*.
[32] S. Kampakis and A. Adamides, "Using Twitter to predict football outcomes," 2014, *arXiv:1411.1243*.
[33] R. Ul Mustafa, M. S. Nawaz, M. I. Ullah Lali, T. Zia, and W. Mehmood, "Predicting the cricket match outcome using crowd opinions on social networks: A comparative study of machine learning methods," *Malaysian J. Comput. Sci.*, vol. 30, no. 1, pp. 63–76, Mar. 2017.
[34] R. Beal, S. E. Middleton, T. J. Norman, and S. D. Ramchurn, "Combining machine learning and human experts to predict match outcomes in football: A baseline model," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 17, pp. 1–5.
[35] J. Pfeffer, D. Matter, and A. Sargsyan, "The half-life of a tweet," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 17, Jun. 2023, pp. 1163–1167.
[36] C. G. Fink, N. Omodt, S. Zinnecker, and G. Sprint, "A congressional Twitter network dataset quantifying pairwise probability of influence," *Data Brief*, vol. 50, Oct. 2023, Art. no. 109521.
[37] C. G. Fink, K. Fullin, G. Gutierrez, N. Omodt, S. Zinnecker, G. Sprint, and S. McCulloch, "A centrality measure for quantifying spread on weighted, directed networks," *Phys. A, Stat. Mech. Appl.*, vol. 626, Sep. 2023, Art. no. 129083.
[38] H. Liu, D. Tam, M. Muqeeth, J. Mohta, T. Huang, M. Bansal, and C. A. Raffel, "Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, Dec. 2022, pp. 1950–1965.
[39] S. Hido, T. Ide, H. Kashima, H. Kubo, and H. Matsuzawa, "Unsupervised change analysis using supervised learning," in *Advances in Knowledge Discovery and Data Mining*. Berlin, Germany: Springer, 2008, pp. 148–159.
[40] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou, "Chain-of-Thought prompting elicits reasoning in large language models," 2022, *arXiv:2201.11903*.
[41] T. Cao, M. Law, and S. Fidler, "A theoretical analysis of the number of shots in few-shot learning," 2019, *arXiv:1909.11722*.
[42] S. Min, X. Lyu, A. Holtzman, M. Artetxe, M. Lewis, H. Hajishirzi, and L. Zettlemoyer, "Rethinking the role of demonstrations: What makes in-context learning work?" 2022, *arXiv:2202.12837*.
[43] J. Sun, Y. Luo, Y. Gong, C. Lin, Y. Shen, J. Guo, and N. Duan, "Enhancing chain-of-thoughts prompting with iterative bootstrapping in large language models," 2023, *arXiv:2304.11657*.
[44] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," 2022, *arXiv:2203.11171*.
[45] J. Liu, A. Liu, X. Lu, S. Welleck, P. West, R. L. Bras, Y. Choi, and H. Hajishirzi, "Generated knowledge prompting for commonsense reasoning," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, S. Muresan, P. Nakov, and A. Villavicencio, Eds. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 3154–3169.
[46] Z. Sun, X. Wang, Y. Tay, Y. Yang, and D. Zhou, "Recitation-augmented language models," 2022, *arXiv:2210.01296*.
[47] D. Zhou, N. Schärli, L. Hou, J. Wei, N. Scales, X. Wang, D. Schuurmans, C. Cui, O. Bousquet, Q. Le, and E. Chi, "Least-to-most prompting enables complex reasoning in large language models," 2022, *arXiv:2205.10625*.

GINA SPRINT (Senior Member, IEEE) received the B.S. degree in computer science from Eastern Washington University, Cheney, WA, USA, in 2012, the Ph.D. degree in computer science from Washington State University, Pullman, WA, USA, in 2016, and the M.B.A. degree from Gonzaga University, Spokane, WA, USA, in 2021.

She was a Faculty Member at Washington State University, from 2016 to 2017. She has been an Associate Professor with Gonzaga University, Spokane, since 2017. She has authored over 30 papers in her interested research areas of machine learning and computer science education.

Dr. Sprint is a member of the Association for Computing Machinery and the IEEE.