

Prediction from Regional Angst – A Study of NFL Sentiment in Twitter Using Technical Stock Market Charting

Robert P. Schumaker¹, Chester S. Labeledz Jr.², A. Tomasz Jarmoszko² and Leonard L. Brown¹

¹Computer Science Dept., University of Texas at Tyler, Tyler, Texas 75799, USA

²Central Connecticut State University, New Britain, Connecticut 06050, USA

rob.schumaker@gmail.com, clabeledz@gmail.com, jarmoszko@ccsu.edu and lbrown@uttyler.edu

Word Count: 7763

Abstract

To predict NFL game outcomes, we examine the application of technical stock market techniques to sentiment gathered from social media. From our analysis we found a \$14.84 average return per sentiment-based wager compared to a \$12.21 average return loss on the entire 256 games of the 2015-2016 regular season if using an odds-only approach. We further noted that wagers on underdogs (i.e., the less favored teams) that exhibit a “golden cross” pattern in sentiment (e.g., the most recent sentiment signal crosses the longer baseline sentiment), netted a \$48.18 return per wager on 41 wagers. These results show promise of cross-domain research and we believe that applying stock market techniques to sports wagering may open an entire new research area.

Keywords: Business Intelligence, Decision Support, Sentiment Analysis, Sports Analytics

1. Introduction

A three-game losing streak and Chicago Bears fans are hopeful, whereas a two-game winning streak and Oakland Raiders fans are expecting the worst. Two different fan-bases and two very different ways of expressing sentiment, one optimistic and the other pessimistic. This type of fan behavior makes the comparison of sentiment between geographically different fan bases a non-trivial problem. In an ideal world with 32 NFL teams, there would be 32 identical fan-bases which would have similar experiences and behave in a fairly consistent, predictable manner. However, this isn't

the case, as illustrated by the earlier example. Each NFL market is a unique and ever-changing mix of individuals with differing perceptions of the world around them. Their perceptions of expected team performance can be captured in their writings (in our case, tweets), analyzed for sentiment and aggregated to form a crowdsourced signal that can be used to forecast the winning team. For example, on a scale of -100%, all negative, to +100%, all positive, the Raiders may have a pre-game tweet sentiment rating of -6.04% and the Bears 36.7%. Following our example further, the Raiders had a record of 2 wins and 1 loss while the Bears had lost all three of their contests. The Raiders were favored in the pre-match betting lines, -179 Raiders to +148 Bears. At that point, did the 36.7% positive tweet sentiment for the Bears have greater predictive value than the teams' record or the betting line? What if Raiders fans are naturally more pessimistic than Bears fans, can we identify a comparable signal in the data? The problem is that making comparisons between teams in an absolute sense and without consideration of their fanbase differences may lead to a less successful prediction. While studies have been successful in using absolute sentiment, which will be discussed in greater detail later [1], [2], we feel that there is a better approach to address this problem.

Our approach is to borrow techniques from technical charting used in stock price analysis. In particular, we analyze sentiment polarity as a time-series signal and examine the position and magnitude of signal change between two temporal windows. In technical charting, a popular technique to analyze price movement is to study the 50 day and 200 day moving averages. Stocks whose 50 day averages cross above their 200 day average are referred to as golden crosses and typify investment opportunity. Whereas stocks whose 50 day averages cross below their 200 day average are referred to as death crosses and typically signal stock price trouble. This type of time-series price signal is similar to that of polarity sentiment prior to a game, which we also can analyze as time-series data.

Polarity sentiment is a measure of the emotional direction of text that is a direct result of author word choice and it can be categorized into three states: positive, negative and neutral. This sentiment data can change quickly over time in reaction to new information, as shown by the Oakland Raiders sentiment in Figure 1, ninety-six hours prior to kickoff.

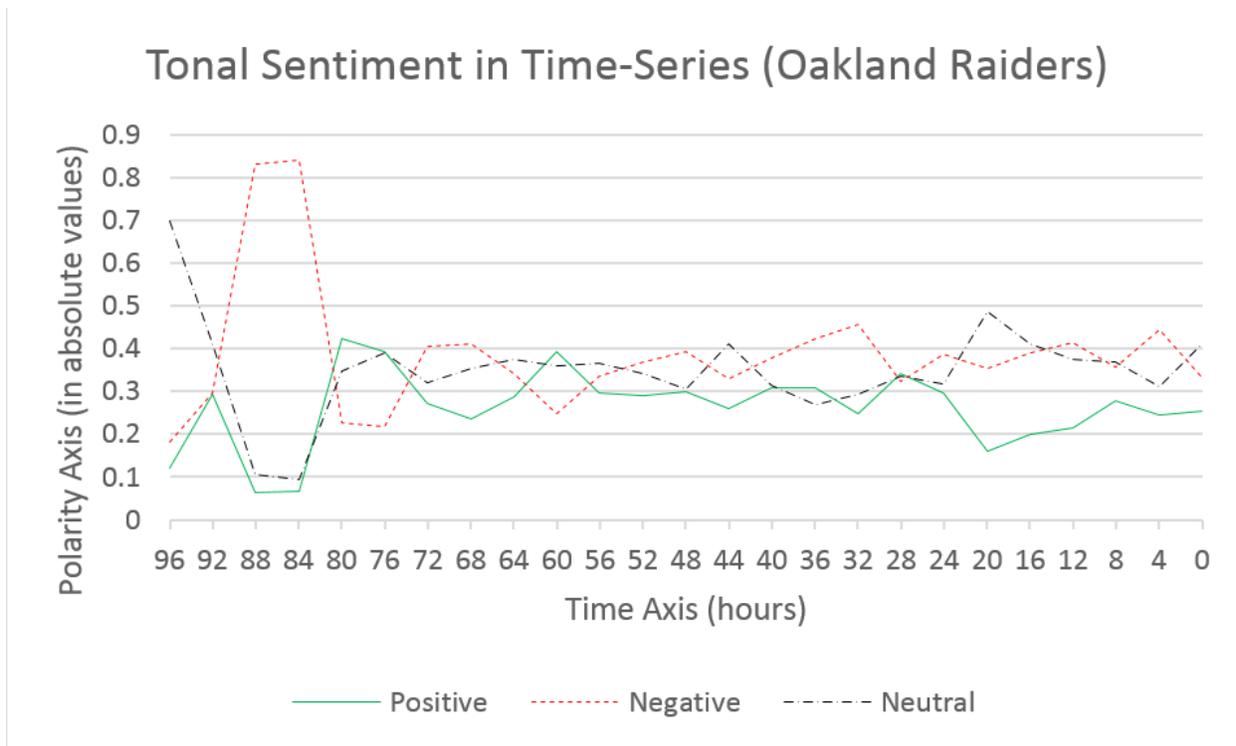


Figure 1. Tonal Sentiment as a Time-Series Signal

Because of the similarities between stock price and sentiment time-series data, we believe that some of the same techniques used to predict stock price movement may also work in sports gaming prediction. By analyzing the aggregated differences of signal between two moving averages, as shown in Table 1, and comparing it between teams, one may avoid problems associated with forecasting accurately despite the noise of NFL fan base differences.

	Sentiment Averages		
Team	96hr Window	24hr Window	Δ Sentiment
Raiders	-41.3%	-60.3%	-19.0%
Bears	-12.2%	-4.1%	8.1%

Table 1. Moving Averages Sentiment Data for Week 4 Raiders and Bears

This is because the moving averages technique in effect normalizes the sentiment data for each team fan-base by capturing the change in sentiments. For this particular example, the Bears have an 8.1% surge in positive sentiment between their 96 hour and 24 hour moving averages and are predicted to win against the Raiders, whom experienced a 19.0% drop in sentiment. In this case the Bears won in a 22-20 upset over bookmaker odds (-179 Raiders, +148 Bears).

While one may argue that using absolute sentiment in either the 96hr or 24hr windows also would have predicted a Bears victory, this is not always the case, as will be later demonstrated. Additionally, in this example we noticed from the data that the Bears exhibited a “golden cross” sentiment behavior where the 24hr sentiment crossed above the 96hr threshold, lending strong predictive support of a Bears win.

The remainder of the paper describes use of sentiment-based technical analysis techniques in more detail. Because both the stock market and sentiment analysis rely on time-series data, we believe that more advanced charting techniques can be borrowed and successfully applied in the future. This paper serves as one of the first steps to examine the crossover possibilities. We further believe that these techniques can be generalized to other sports in which predictions from differing fan bases may be of interest.

The rest of this paper is organized as follows. Section 2 investigates text analytics (including sentiment analysis), wagering and crowdsourcing, and technical charting. Section 3 contains our research questions. Section 4 introduces the CentralSport system, a fusion of sentiment analysis and technical stock market charting. Section 5 presents our experimental design. Section 6 contains the

experimental results and a discussion of their implications. Section 7 concludes with a summary of findings and future research directions.

2. Literature Review

Social media sentiment analysis has been shown useful in a wide area of prediction domains. However, it can only provide a snapshot of author activity and is only one such tool in the text mining repertoire. If we wish to apply meaning and decision-making to these snapshots of information, we must look to a higher level of abstraction, text analytics.

The rest of this section focuses on several key foundations; text analytics, where we examine many of the different ways text information can be extracted from text data and transformed into knowledge, including sentiment; wagering and crowdsourcing, where we explore the wisdom of the crowds and how it relates to wagering activity, and technical charting techniques as it relates to stock prediction.

2.1 Text Analytics

Text Analytics can be defined as a process of extracting information from unstructured text, typically using some form of Natural Language Processing (NLP), that can be analyzed for patterns, trends and tendencies and applied to business intelligence problems [3]. This field differs from Information Retrieval (IR), which is concerned with finding relevant documents. Instead, text analytics focuses on representation prior to delivery to a user, as shown in Figure 2.

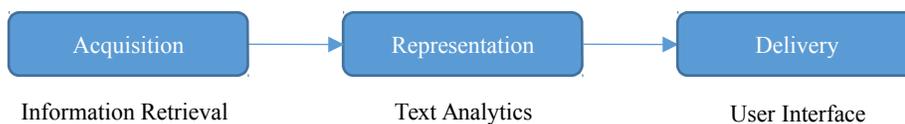


Figure 2. Acquisition, Representation and Delivery of Text

Text analytics differs from text mining in several ways. If we were to examine texts from the perspective of the DIKW hierarchy [4], text analytics resides at the knowledge creation stage and uses tools to extract a deeper meaning from text, such as understanding context and extrapolating semantic meaning to provide a fuller sense of the material. These tools interact with text mining, which can be defined as extraction-only techniques [4] at the Information layer and by themselves cannot solve business problems. Thus text mining focuses on a lower part (e.g., information) of the DIKW hierarchy, while text analytics focuses on knowledge creation, as shown in the text analytics taxonomy of Figure 3. Whereas understanding context and semantic meaning cross knowledge and information.

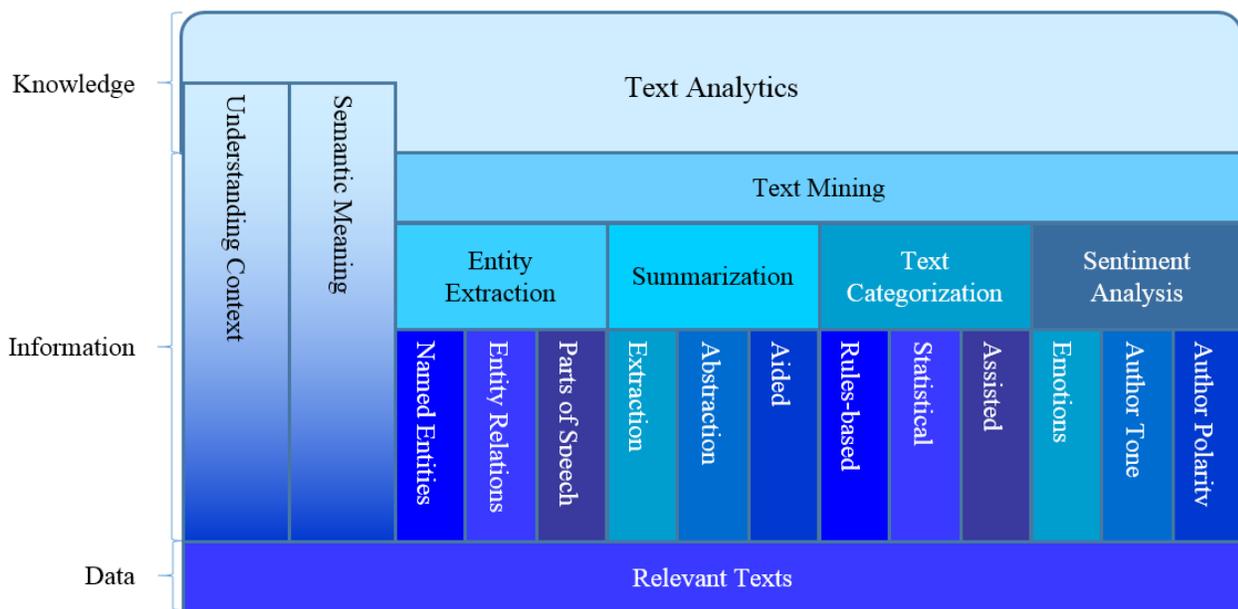


Figure 3. A Text Analytics Taxonomy in Relation to the DIKW Hierarchy

Text mining encompasses several distinct areas: entity extraction, summarization, text categorization and sentiment analysis [3], [4]. Each of these areas will be explained in more detail throughout the following sections.

2.1.1 Entity Extraction

Entity Extraction is the process of recognizing specific instances within a text, such as named entities, entity relations and parts of speech components [4]. In named entities, the goal is to extract specific information from an unstructured or semi-structured text. An example business application would be a type of event extraction system [5] that could scan an individual's email for flight confirmations, extract the times and airports involved, and automatically enter it into a personal calendar.

Entity relations aims to find textual relationships among two or more entities. One example is discovering business relationships within financial documents in order to build a social network graph of individuals and their contacts.

Parts of speech (POS) extraction is a field of computational linguistics that focuses on the identification of component grammar entities that can be used as a type of representation for further analysis in text mining or summarization. These representations can take many forms, however, the most popular are bag of words, followed by proper nouns and named entities POS.

In a bag of words approach, text is filtered to remove semantically meaningless “stopwords” such as [a, the, of]. Then, typically, word stemming is used to identify root words (e.g., “stopping” and “stopped” become “stop”). The remaining text is then used for representation. This standard of text representation is used extensively within the financial sector [6] and can be adapted into rules [7]. Noun phrase representation uses only the nouns and noun phrase segments of a sentence for representation. This representation is much lighter than bag of words and has been found to adequately convey critical document ideas [8]. Proper nouns are a subset of noun phrases in which only nouns of a specific nature (e.g., names of individuals or businesses) are retained for representation. While much more constraining and lighter than noun phrases, proper nouns was found to be a better representation

scheme for financial text than other representations [9]. Named entities POS is an even more restrictive noun phrase subset, in which only the nouns that conform to the categories of date, location, money, organization, percentage, person and time are used. This form of text representation was found to be too restrictive with respect to financial text because of the fixed categories [9].

2.1.2 Summarization

The field of text summarization is tasked with condensing the salient concepts of a text into a shorter version. This shorter version could be anywhere from a couple of sentences to several pages in length, depending on requirements. Text summarization can be approached in several ways, including extraction, abstraction and human-aided. Extraction-based approaches form the bulk of text summarization because of their relative ease with current technologies. These approaches use original content and must decide how to reduce, condense or abridge the source document [10].

Abstraction-based approaches generally require a level of understanding of a text before generating a response. This typically requires the use of NLP, lexicons and domain-specific ontologies [10].

Aided approaches involve human intervention and feedback. In this approach, several summaries may be generated and the human participant either selects the best one, or may choose to have the next summarized iteration focus on different keywords [11].

2.1.3 Text Categorization

Text categorization is the process of matching a previously unseen text against predefined categories. This process could be rules-based, in which specific terms or patterns are tested (e.g., simple spam filters), statistical, where the likelihood of membership in a category is computed based on keywords (e.g., topic classification), or assisted classification, in which a human domain expert may intervene to modify classification results [12].

2.1.4 Sentiment Analysis

Sentiment analysis is the science of identifying opinions and emotions in text [13], which can include author tone, whether a text is objective or subjective, and author polarity, whether the text is positive or negative [14]. These features are embedded within a text and are a function of author word choice, which can often provide insight into an author's emotional state.

Tracking sentiment through social media has been shown useful to predict elections, stock markets and sports. The 2009 German federal elections were examined using a Twitter volume count for each of the candidates as a proxy for voting intention. This simple technique had a mean average error of 1.65% with positive sentiment a leading indicator [15]. While the results obtained were good for a direct election democracy like Germany, critics would argue that elections that use a form of proportional representation (e.g., electoral college), sentiment measures at the regional, jurisdictional level would be better predictors.

Sentiment has also been successfully used to predict stock price movement, using the word choices of breaking financial news articles. One study investigated near-term stock prices and found that sentiment was a better indicator of stock price increase (3.04% trading return) than using a baseline S&P500 index with a 2.41% trading return during the same period [16].

The forecasting of sport has an avid following, ranging from fantasy team owners to professional gamblers and even academics. In a study of NFL game prediction, Sinha et. al. used Twitter features, such as changes in tweet volume, to build a model capable of 55% accuracy, exceeding the breakeven point of a bookmaker's commission [1]. In a study of sentiment on the prediction of English Premier League soccer results, Schumaker et. al. examined aggregated tweet sentiment in the 96 hours prior to match start between the two clubs. They found that using sentiment netted a higher payout, \$2,704.63, compared with the return loss (\$1,887.88) of an odds-only approach

[2]. They also observed that contrasting a club's weekly tweet difference with its seasonal average netted an even greater payout, \$3,011.20. This examination of volume change versus club average was a step towards minimizing fanbase differences and shows the promise of further research in this direction.

A variety of methods can be used to extract sentiment [17]. One well-known tool to measure sentiment is OpinionFinder, a tool which can identify author tone and polarity in text [14]. When compared against the MPQA Opinion Corpus, OpinionFinder had an accuracy of 74%, subjective precision of 78.4%, subjective recall of 73.2% and a subjective F-measure of 75.7%, as compared to baseline accuracy of 55.3%. While this tool was originally created with the intended use on longer texts which can contain additional information [18], it was found to be useful on the much shorter Twitter tweets as well [2].

2.2 Wagering and Crowdsourcing

Sports gaming and the stock market share a lot of similarities. Both handle large amounts of currency and both allow individuals to forecast future events. The worldwide sports gaming market is worth an estimated \$1 trillion dollars whereas the New York Stock Exchange (NYSE) is around \$16 trillion [19]. Both markets also have long-term participants that typically use research and skill before committing their resources to a forecast. There are also many differences; for example, stock markets face much heavier government regulation. A case in point is how these markets deal with "insider trading" (e.g., the use of non-public information for personal gain). In the United States, statutes made insider trading illegal in the wake of the 1929 stock market crash, but those same laws do not extend to US sports gaming. One recent example was in 2015 where several employees from DraftKings used private company data to make millions of dollars on their rival site FanDuel [20]. While this case could not be prosecuted as insider trading because sports gaming does not fall within the legal

definition of a financial market, there are other legal challenges claiming unfair business practices yet to be decided by the courts. Another sports gaming example that could be viewed as a form of insider trading is courtsiding. In this activity a bettor will place bets on in-game events (e.g., a fault that just occurred in a tennis match) prior to the gaming markets reflecting that information [21]. This activity is actively discouraged in professional tennis, but could just as easily be ported to other sports. While courtsiding faces some ethical challenges, profits can be made with superior data and fast access to markets.

Wagering on the outcome of U.S. professional football games dates back more than a hundred years [22]. In the early 1900s, bettors simply tried to predict winners without regard to their margin of victory. Beginning in mid-century “point spread” betting, wagering based on the final score differential, became a central fixture in North American football wagering. The goal of the spread is to balance the attractiveness of wagering between competing teams, even if they seem mismatched. To win at the spread the favored team must either meet or exceed the score differential, whereas the non-favored team must either 1. win or 2. lose by fewer points than the spread.

Spread wagering is also similar to how modern bookmakers set odds. Before two teams meet on the field, bookmakers set initial betting lines (i.e., odds) in an attempt to draw an equal currency amount of wagers on each team. Their goal is to balance the wagers, so that the losing bettors pay the winners, minus the bookmakers’ commission. It is in the bookmakers’ best interests to periodically adjust odds and rebalance the line to avoid the potential of monetary loss. Therefore a bookmaker will increase odds on a less-wagered team to give bettors an incentive to increase betting on it and thus rebalance the line.

One type of popular wagering system is the Moneyline. In this system, teams with negative values are favored and teams with high positive values are longshots. Odds and payouts (i.e., return on

successful wagers) are based on a unit of \$100. For example, the New York Giants and Cleveland Browns may have a Moneyline of -250 and +500 respectively. For the bettor on the Giants (the favorite) they would need to wager \$250 to win \$100. For the Browns bettor, they would wager \$100 in a bid to win \$500.

Bookmakers set initial odds by using historical wagering data and intuition. Once those odds are set, the bookmakers will periodically rebalance the wagers between teams. Because of the number of bookmakers available to accept wagers, wagering imbalances may develop and create odds differences between books. Typically, those bookmakers with the more favorable odds will attract more wagering, which will force their odds to return to market equilibrium.

Crowdsourcing offers one way to generate accurate and reliable forecasts that can be used for wagering. This technique uses an average of individual forecasts to predict future events [23]. In a study of the 2006 FIFA World Cup soccer matches, crowds were found better able to predict winners than were comparative FIFA national team rankings or random chance [24]. In a study of German Premier League soccer, the wisdom of crowds was found to be more accurate than bookies [25]. In a study of English Premier League soccer, social media sentiment (a type of crowdsourced data) was similarly accurate [2].

2.3 Technical Charting

Stock prices can be thought of as time-series data that is difficult to discern patterns from in the near-term, but easier to forecast in the long-term. This signal will adjust with the introduction of new information into the market, regardless of whether that information is public or private. Financial news articles are one such source of public information that, depending on the information content, can provoke differing price responses [26]. This desire for new forms of market information has transformed how brokerage houses approached securities trading [27].

With regards to stock price forecasting, we must acknowledge the two major tenets of securities pricing strategies, fundamental and technical analysis. In fundamental analysis, the underlying metrics of company fiscal health (e.g., ROA, ROE, D/E, P/E, etc.) are examined to determine price forecasts that are generally of a longer duration (i.e., 6 months or longer). Conversely, technical analysis examines price and/or volume signals and compares that data to historical values in an attempt to identify known activity patterns. These strategies are of a much shorter duration and are a favorite of market daytraders.

Technical Analysis is not without its critics whom point out a lack of consistent empirical data to support the claim that stock prices could be predicted based on “fit” to some historical chart [28], [29]. However, technical analysis has many proponents whom argue that studies have validated their predictions of stock price movements [30], especially for short-term timeframes and index/industry specific contexts [31]. Although this disagreement has yet to be resolved, technical analysis techniques have been gaining in popularity.

One such technique is moving averages, which helps to lessen daily price volatility by providing a moving window of averaged historical prices which functions as a stabilizing weight when trying to discern a trend. Typically, two or more moving averages are computed and the interaction between them becomes of interest [32]. When these time-series signals cross one another, it typifies a buy/sell recommendation.

There have been several studies in the finance literature using moving averages. In a study that used twenty-one years worth of data from the Singapore Stock Exchange, researchers concluded that investment decisions based on moving averages could generate a significant positive financial return [33]. Another study examined a hundred years of Dow Jones Industrial Average data and found support for technical analysis techniques [34]. A third study used technical analysis techniques coupled

with portfolio building and found that a one-week formation period using a momentum strategy netted a 20.79% trading return over a five-week holding period [35].

2.4 Research Gaps

Our review of the literature identified several opportunities not previously pursued, notably very few sentiment studies on sports wagering. We seek to extend the existing research by investigating polarity measures from a technical charting standpoint and determine if the application of finance techniques can be successfully applied to improve betting returns.

Another gap was that prior studies treated the sentiment of each team equally, ignoring fan-base differences. Our work not only takes these differences into account but also allows an inspection of sudden, recent changes in sentiment. Our intent is to derive a profitable system based on crowdsourced information that is not typically used in a wagering market.

3. Research Questions

To address the research gaps and explore this area further, we propose the following research questions with a brief explanation to follow.

1. Can the technical charting of fan sentiment predict game outcomes?

From prior studies, counts and contents of tweets and blogs appear to function as adequate proxies for prediction. Those studies all rely on one set of time-series data which ignores fan-base differences and in turn obscure the data signal. By teasing out a baseline of fan sentiment and only tracking the changes, we suspect to find a more predictable model.

2. Can fan sentiment be profitable?

Prior studies in other sports domains have shown an inverse relationship between accuracy and profitability. Does the same hold true in the NFL and if so, can fan sentiment be used for profitable wagering?

3. *What technical charting patterns can lead to improved wagering?*

Time-series sentiment can provide clues as to expected team performance that may not be reflected in bookmakers' odds. Can signals within the data, similar to stock market price resistance and support levels, be used to indicate sudden changes in expected team performance and improve bettors' results?

4. System Design

Our approach is to test modified stock price charting techniques used by Wall Street to represent sentiment signals relative to regional affect. In technical charting two moving averages (50 and 200 day) are typically utilized. The idea is that the longer moving average provides a type of baseline behavior and the shorter moving average represents recent activity but is long enough to smooth out any short-term volatility. As applied to the stock market, when the 50 day moving average crosses above the 200 day support level (a golden cross), technical market analysts view the stock favorably and it likely posts price gains. When the 50 day moving average crosses below the 200 day resistance level (a death cross), this signals excessive pessimism about the investment. Applied to our study of NFL Twitter sentiment, we chose 96 hour and 24 hour periods prior to game start, which are sufficient to establish a weekly game baseline and smooth out any volatility.

Our system is illustrated in Figure 4 with an explanation of critical components to follow.

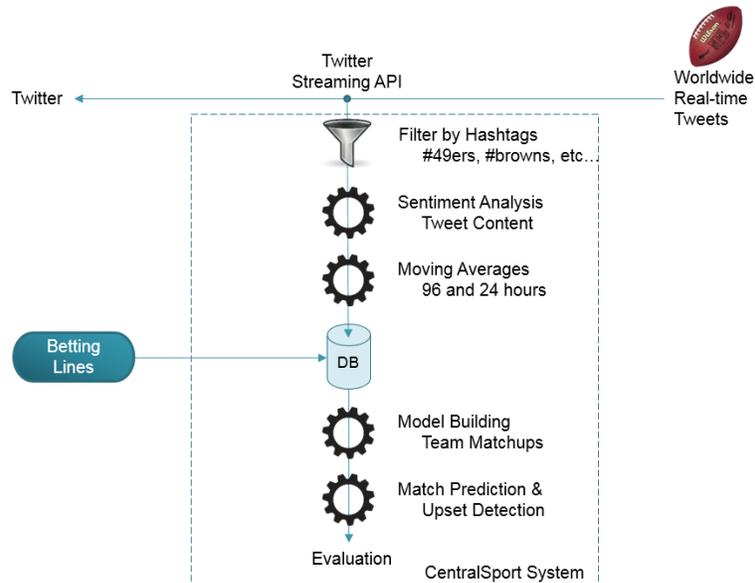


Figure 4. The CentralSport System

Tweets are gathered in real-time from the Twitter Streaming API and filtered by hashtag for each of the 32 NFL teams. To limit the effect any one tweeter could have on the results, we only analyzed one tweet per tweeter in each of the moving average windows; that is, if a tweet author created more than one message we only used the first one. While we recognize that this could be a study limitation, we sought to prevent a minority of prolific tweeters from skewing the results. The content of each tweet is then passed through OpinionFinder to determine tweet polarity: positive, negative or neutral. Since OpinionFinder works at identifying sentiment at the sentence-level and that some tweets may contain multiple sentences, we used a majority-rules approach in determining tweet polarity. In cases where the number of positive and negative values were equivalent, those tweets were marked neutral. Moving averages for 96 hours and 24 hours prior to kickoff are then computed for the sentiment polarity values. Betting lines at the start of each game are then collected from OddsPortal.com, a compilation of several wagering firms that are averaged using the Moneyline approach.

Four models are then built to test regional sentiment values. The first is a baseline odds model that wagers on favorites using OddsPortal.com data. The second model uses the sentiment gathered 96 hours prior to kickoff. The team with the higher normalized positive to negative sentiment ratio is predicted to win, as shown in Equation 1.

$$Sentiment_{Team,96hr} = \frac{Sentiment_{Team,96hr,Pos}}{Sentiment_{Team,96hr,Total}} - \frac{Sentiment_{Team,96hr,Neg}}{Sentiment_{Team,96hr,Total}} \quad (\text{Equation 1})$$

The third model is a variation of the second and uses the sentiment gathered 24 hours prior to kickoff. Both the 96 hour and 24 hour models ignore fan-base differences. These models represent current state-of-the-art sentiment analysis as it relates to time-series data and are also important for comparison purposes.

The fourth model uses technical charting and takes into account the change in sentiment during these periods and calculates the sentiment surge or drop (we refer to it as swing) according to Equation 2 for each team prior to game start.

$$Swing_{Team} = Sentiment_{Team,24hr} - Sentiment_{Team,96hr} \quad (\text{Equation 2})$$

Swing values for each team are then compared and the team with the higher swing value is predicted to win. We then evaluate each model’s predictions against actual outcomes and calculate accuracy and payout values. Payout uses a simulated trading engine that wagers \$100 on each prediction and then calculates the theoretical winnings (or losses) using a Moneyline odds calculator from <http://www.breakingodds.com/page/calculators>. Betting efficiency is calculated as the payout divided by the number of wagers. An example of the system using the Steelers and the Patriots from Week 1 of the 2015-2016 NFL season is shown in Table 2.

	96hour Sentiment			24hr Sentiment		
	Positive	Negative	Neutral	Positive	Negative	Neutral
Steelers	6,400	3,696	2,893	5,214	2,575	2,436
Patriots	31,510	7,126	2,178	21,088	1,265	2,094

Table 2. Sentiment values for Steelers vs Patriots Sept. 10, 2015

Running these values through Equations 1 and 2 results in the values in Table 3.

	Sentiment Models				Score
	Baseline	96hr	24hr	Swing	
Steelers	272	20.82%	25.81%	4.99%	21
Patriots	-345	59.74%	81.08%	21.34%	28

Table 3. Prediction Results of Steelers vs Patriots Sept. 10, 2015

From Table 3, all four models predicted the Patriots to win. Baseline, using the odds-only approach from OddsPortal.com, chose the lower odds Patriots (-345 to +272). The three sentiment models also picked the Patriots because of the higher positive sentiment. Focusing on the swing model, the Steelers' swing value is 4.99% indicating that, after normalizing the data, the 24 hour sentiment was 4.99% more positive than the 96 hour sentiment. Fans were expressing more optimism in the hours leading up to kickoff. Additionally, the Patriots' swing value was 21.34% indicating greater twitter enthusiasm. All four models correctly picked the Patriots to win, which they did by a score of 28-21.

5. Experimental Design

For the experiment we chose the 2015-2016 NFL regular season as our testbed. This dataset encompassed 17 weeks and a total of 256 games. During this period we collected tweets in real-time using one hashtag identified by a domain expert for each of the 32 teams. Following prior work we used the 96 hours prior to each game's kickoff for our study, acknowledging that some teams had Sunday followed by Thursday games which may cause a slightly more impure dataset than we would have liked. This 96hr window became the long-term sentiment signal. Drawing inspiration from technical charting, where there exists a factor of four difference between long and short-term window sizes (e.g. 200 day vs 50day moving averages), we chose 24 hours as our short-term horizon. From this process, 4,509,260 tweets were used in the study. Week 12 (Thanksgiving week) had the most tweets at 345,218 and Week 4 (the last regular season week of Major League Baseball) had the fewest with

142,760. The weekly average number of tweets was 265,251 with a standard deviation of 60,662. The pattern of weekly tweets is shown in Figure 5.

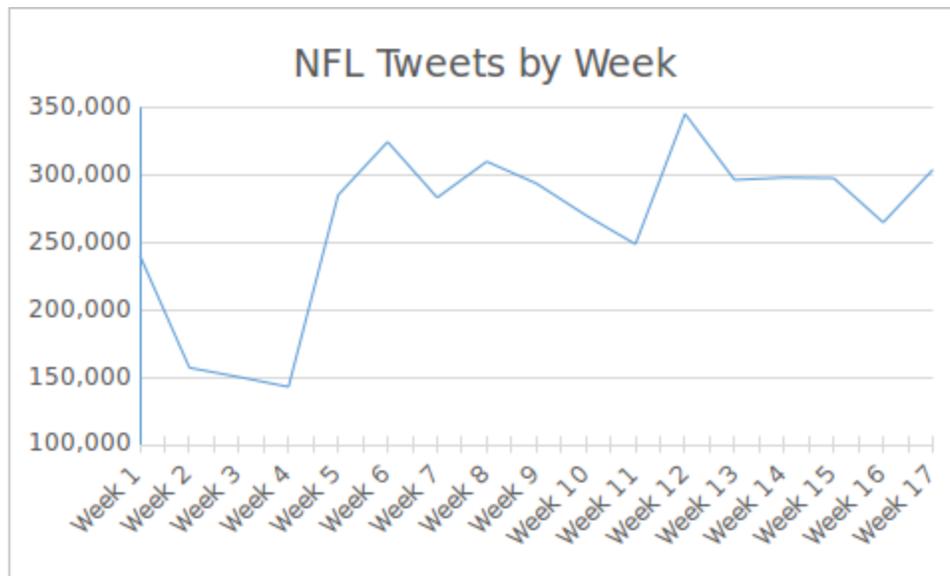


Figure 5. NFL Tweets by Week

The New England Patriots had the most tweets of any team with (315,525) and the New York Giants had the fewest at 28,238 because the hashtag used, #nygiants, was selected to not interfere with the San Francisco Giants baseball team. While the differences in tweet counts may prompt questions, we noted similar differences in prior research and took care to normalize the NFL data to minimize these effects. The average team number of tweets was 140,914 with a standard deviation of 69,331. Figure 6 shows the tweets collected for each team.

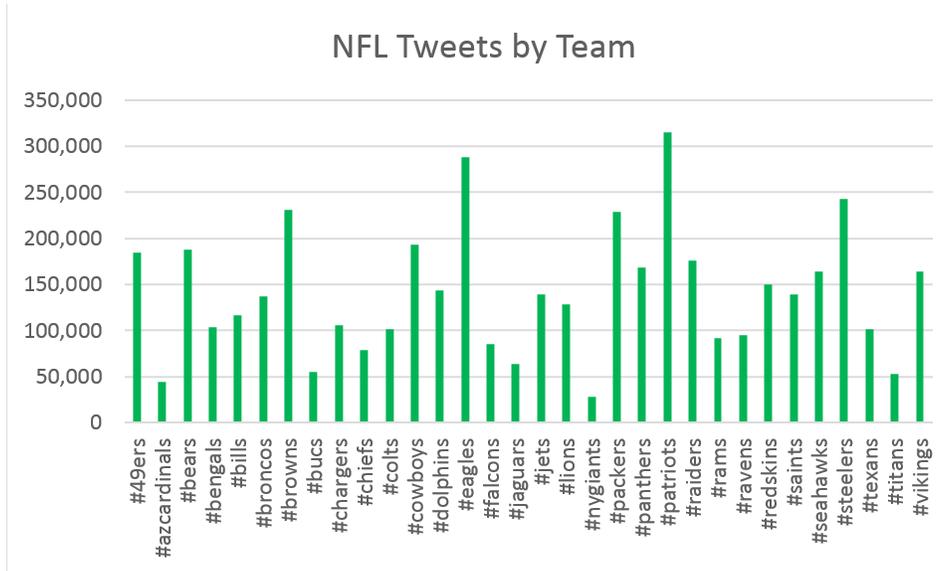


Figure 6. NFL Tweets by Team

6. Experimental Results and Discussion

6.1 Technical Charting of Fan Sentiment and Accuracy

To answer our first research question of *can the technical charting of fan sentiment predict game outcomes* we made accuracy predictions on all 256 NFL games during the 2015-2016 regular season, comparing the four models against the actual outcome. The results are presented in Table 4.

	Baseline	96hr	24hr	Swing
#Correct	153	128	137	150
Average	59.8%	50.0%	53.5%	58.6%
p-value		0.013	0.077	0.394

Table 4. Accuracy Predictions of the Models

From this table, using the odds-only approach, Baseline, achieved an accuracy of 59.8% by correctly predicting 153 of 256 games. The model using the Twitter sentiment 96 hours prior to kickoff was accurate 128 times for 50.0% accuracy. The 24 hour model was slightly more accurate at 53.5%, correct on 137 games. Swing, the fusing of the two prior models into a technical charting technique, was accurate 150 times or 58.6%. From these values, Baseline and Swing were statistically equivalent, $p\text{-value} = 0.394$. Comparing Baseline accuracy to the 96 hour and 24 hour models independently, the sentiment models alone were found to perform statistically worse. This result shows that using

absolute sentiment by itself in either a 96hr or 24hr timeframe does not have better accuracy than odds, which is consistent with the accuracy results of absolute sentiment comparisons in the prior English Premier League study. However, we would point out that these two sentiment models had accuracy of 50% or more indicating performance at or slightly in excess of random chance. This slightly depressed value may be attributable to signal noise.

Diving deeper into the results and looking at predictions on favorites versus upsets, the Swing model selected favorites 54.3% of the time on 139 games, and upsets 45.7%, as shown in Table 5. It was interesting to note that all models were more correct versus incorrect when predicting favorites to win and more incorrect than correct in predicting upsets.

Pattern	Other Team's Pattern	Wagers	Accuracy	Payout	Efficiency
Parallel Upwards	Parallel Upwards	27	70.4%	\$949.67	\$35.17
Parallel Upwards	Parallel Downwards	39	56.4%	\$210.91	\$5.41
Parallel Upwards	Crossing Upwards	37	56.8%	\$1,013.24	\$27.38
Parallel Upwards	Crossing Downwards	24	66.7%	\$469.31	\$19.55

Table 5. Model Accuracy on Favorites and Upsets

From this data, when Swing wagered on the odds favorite team, the accuracy was 66.9% (93 to 46) versus wagering on upsets with 48.7% accuracy. Comparing Baseline to Swing on Favorites Accuracy, 59.8% to 66.9% respectively, we noted a p-value of 0.05 indicating that Swing had a statistically significant improvement in accuracy. However, each absolute sentiment model (96hr and 24hr) for Favorites, yielded 59.3% and 62.9% accuracy respectively. Neither of which was found to be statistically different from Baseline accuracy.

It would appear that a wagering strategy of selecting those teams that are both odds favorites and have a more positive Swing value can yield better accuracy than odds alone. From the data, neither the 96 hour nor 24 hour sentiment models had statistically different results than the odds-only Baseline approach on favorites. This would indicate that the technical charting method of the Swing model is

adding value. This finding can have implications on betting houses by exposing a technique for improving wagering odds, and bettors for improving their wagering accuracy.

6.2 Technical Charting of Fan Sentiment and Profitability

To answer our second research question of *can fan sentiment be profitable* we made hypothetical \$100 wagers on matches and analyzed their returns for each model, as shown in Table 6.

	Baseline	96hr	24hr	Swing
Sum	(\$3,125.79)	(\$545.10)	\$1,237.91	\$3,798.15
Average	(\$12.21)	(\$2.13)	\$4.84	\$14.84
p-value		0.115	0.022	0.0009

Table 6. Payout Predictions of the Models (parentheses indicate negative numbers)

From this table, the odds-only Baseline had a return of -\$3,125.79 for an average loss of \$12.21 per wager. Given the 59.8% accuracy from earlier coupled with the lower odds for favorites, the accuracy was not sufficient to offset losses which amounted to \$10,300 versus \$7,174.21 in winnings. For the absolute sentiment models, the 96 hour model had a payout loss of \$545.10 and the 24 hour model demonstrated a payout gain of \$1,237.91. These two models coupled with their lower accuracies from earlier indicate that crowdsourced sentiment was identifying longshot opportunities, which is in line with prior studies. The technical charting model Swing had a payout gain of \$3,798.15 or an average of \$14.84 profit per wager. Swing was found statistically equivalent in terms of accuracy, however, the total payouts from longshot predictions is much improved (p-value = 0.0009).

Looking further into longshot predictions, we examine the payout differences between wagering favorites and upsets as shown in Table 7.

	Favorites		Upsets	
	Correct	Incorrect	Correct	Incorrect
Baseline	\$7,174.21	(\$10,300.00)		
96hr	\$3,894.89	(\$5,500.00)	\$8,360.01	(\$5,400.00)
24hr	\$3,950.32	(\$4,900.00)	\$9,187.59	(\$7,000.00)
Swing	\$4,270.32	(\$4,600.00)	\$10,127.83	(\$6,000.00)

Table 7. Model Payouts on Favorites and Upsets

From this data, none of the models wagering on favorites posted gains which was not surprising given the inherently lower odds. Despite accuracies of 59.8%, 59.3%, 62.9% and 66.9% for each of the four models respectively, none managed a positive payout. However, turning attention towards predicting upsets, the three sentiment models posted gains of \$2,960.01, \$2,187.59 and \$4,127.83 respectively on 102, 124 and 117 wagers. This translates into an average return per wager of \$28.73, \$17.64 and \$35.28 per model respectively on upsets, despite accuracies of 47.1%, 43.5% and 48.7%. We acknowledge that these wagers are hypothetical and that real wagers might influence the betting line, albeit marginally, for slightly lower returns.

These results provide strong evidence that technical charting of social media sentiment can be used profitably in sports gaming. We believe that the longer 96 hour window provides a sort of baseline of sentiments that accounts for fanbase differences when measured against a 24 hour window. We believe that this method provides additional information especially in identifying upsets.

6.3 Technical Charting Patterns

To answer our third research question of *what technical charting patterns can lead to improved wagering* we examined results from four well known technical charting patterns; parallel upwards, parallel downwards, crossing upwards (a.k.a. golden cross) and crossing downwards (a.k.a. death cross) as shown in Table 8.

Pattern	Wagers	Accuracy	Payout	Efficiency
Parallel Upwards	127	61.4%	\$2,643.13	\$20.81
Parallel Downwards	110	61.8%	\$1,731.61	\$15.74
Crossing Upwards	95	55.8%	\$1,484.57	\$15.63
Crossing Downwards	110	56.4%	\$1,060.67	\$9.64

Table 8. Results by Technical Charting Patterns

From this data, if one team exhibited a parallel upwards pattern (i.e., both the 24 hour and 96 hour sentiment moving average were gaining positive values without crossing) the wagering efficiency was found to have an excess return of \$20.81 per wager with 61.4% accuracy. This means that their team sentiment steadily improves prior to kickoff and winners can be predicted 61.4% of the time. Recall from earlier that a Baseline odds-only approach had 59.8% accuracy. Diving further into the observed parallel upwards pattern, we examine its results in matchups with the other team's patterns, as shown in Table 9.

Pattern	Other Team's Pattern	Wagers	Accuracy	Payout	Efficiency
Parallel Upwards	Parallel Upwards	27	70.4%	\$949.67	\$35.17
Parallel Upwards	Parallel Downwards	39	56.4%	\$210.91	\$5.41
Parallel Upwards	Crossing Upwards	37	56.8%	\$1,013.24	\$27.38
Parallel Upwards	Crossing Downwards	24	66.7%	\$469.31	\$19.55

Table 9. Parallel Upwards Results by Technical Charting Patterns

From this data, when both teams exhibited a parallel upwards pattern, the wagering efficiency was \$35.17 with 70.4% accuracy. This means that both teams experienced steadily improved positive sentiment, however, the team with the greater improvement wins 70.4% of the time. If we examine this pattern further, for favorites wagering it is 85.7% accurate on 14 attempts with a \$404.59 payout. For upsets wagering, dual parallel upwards is 53.8% accurate on 13 attempts with a \$545.08 payout.

If we focus solely on upsets, we would expect crossing patterns to be better predictors. Our reasoning is that sudden changes in sentiment might not be captured in the wagering odds and it creates an arbitrage opportunity, as shown in Table 10.

Pattern	Wagers	Accuracy	Payout	Efficiency
Parallel Upwards	55	49.1%	\$2,182.46	\$39.68
Parallel Downwards	46	52.2%	\$1,757.87	\$38.21
Crossing Upwards	41	46.3%	\$1,975.28	\$48.18
Crossing Downwards	58	48.3%	\$1,548.47	\$26.70

Table 10. Upset Results by Technical Charting Patterns

From this table we note that crossing upwards has the highest wagering return of \$48.18 and crossing downwards the lowest at \$26.70, consistent with expectations. For teams crossing upwards a sudden swing in sentiment led to better wagering returns. Likewise, crossing downwards indicates recent team pessimism and might indicate a pending loss (e.g., lower wagering efficiency). If we were to restate data by only wagering based on crossing patterns, e.g., bet on crossing upwards to win, bet on crossing downwards to lose and no bet when both teams show crossing patterns (regardless of favorites or upsets), we would have the following. Crossing upwards would have an accuracy of 54.2% on 59 wagers for \$868.88 or a \$14.73 return per wager. Conversely if we were to wager against teams with a crossing downwards pattern we would have a 54.8% accuracy on 62 wagers for -\$333.61 or a -\$5.38 return per wager. It would appear that wagering on patterns can net a profit and that there is information contained within sentiment crossing patterns as evidenced by their returns.

It was also noted that sometimes a pattern indicated new information that may not have been reflected in the odds. One such example was found in the collapse of the Oakland Raiders on October 4, 2015 as shown in Figure 7.

both the 96hr (50.0% accuracy) and 24hr (53.5% accuracy) timelines. If we were to examine predictions of just favorites, Swing's accuracy increased to 66.9%, a 7.1% higher accuracy value than the odds-only approach which is more than sufficient to offset a bookmaker's commission. In terms of payouts, Swing wagering netted a hypothetical return of \$3,798.15 or \$14.84 per wager versus a \$12.21 loss per wager on odds favorites. For predictions of upsets, Swing again outperformed with an average return of \$35.28.

Looking deeper into the technical charting patterns of team time-series sentiment, we found that if one team exhibited a parallel upwards trend, the average wager return was \$20.81. If both teams exhibited parallel upwards behavior, the team with the greater sentiment improvement won 70.4% of the time with an average wager return of \$35.17. If we focus solely on upset predictions with parallel upward behavior, average wagering returns were \$39.68, and crossing upwards trends (i.e., golden crosses) netted an average \$48.18 return per wager.

While these results are interesting, the idea of porting techniques from technical analysis to sports wagering is promising. These results may be generalizable to other sports domains and the examination of further stock price prediction techniques.

References

- [1] S. Sinha, C. Dyer, K. Gimpel, and N. A. Smith, "Predicting the NFL using Twitter," presented at the ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics, 2013.
- [2] R. P. Schumaker, A. T. Jarmoszko, and C. Labeledz Jr., "Predicting Wins and Spread in the Premier League Using a Sentiment Analysis of Twitter," *Decision Support Systems*, vol. 88, pp. 76–84, 2016.
- [3] Predictive Analytics Today, "What is Text Analytics?," 2016. Retrieved September 10, 2016 from <http://www.predictiveanalyticstoday.com/text-analytics>.

- [4] T. Reamy, *Deep Text: Using Text Analytics to Conquer Information Overload, Get Real Value From Social Media, and Add Big(ger) Text to Big Data*. Medford, New Jersey: Information Today, Inc., 2016.
- [5] F. Hogenboom, F. Frasincar, U. Kaymak, F. de Jong, and E. Caron, “A Survey of event extraction methods from text for decision support systems,” *Decision Support Systems*, vol. 85, pp. 12–22, 2016.
- [6] R. P. Schumaker and H. Chen, “Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System,” *ACM Transactions on Information Systems*, vol. 27, 2008.
- [7] W. Nuji, V. Milea, F. Hogenboom, F. Frasincar, and U. Kaymak, “An Automated Framework for Incorporating News into Stock Trading Strategies,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 4, pp. 823–835, 2014.
- [8] K. M. Tolle and Hsinchun Chen, “Comparing Noun Phrasing Techniques for Use with Medical Digital Library Tools,” *Journal of the American Society for Information Science*, vol. 51, pp. 352–370, 2000.
- [9] R. P. Schumaker and H. Chen, “A Quantitative Stock Prediction System Based on Financial News,” *Information Processing and Management*, vol. 45, pp. 571–583, 2009.
- [10] Udo Hahn and Inderjeet Mani, “The Challenges of Automatic Summarization,” *IEEE Computer*, vol. 33, pp. 29–36, 2000.
- [11] C. Prakash and A. Shukla, “Human Aided Text Summarizer ‘SAAR’ Using Reinforcement Learning,” presented at the 2014 International Conference on Soft Computing & Machine Intelligence (ISCMCI), 2014.
- [12] Stephanie Lemieux, “Auto-Classification: Friend or Foe of Taxonomy Management?,” 2012. Retrieved September 11, 2016 from <http://www.cmswire.com/cms/information-management/autoclassification-friend-or-foe-of-taxonomy-management-014222.php>.
- [13] Ahmed Abbasi, Hsinchun Chen, and Arab Salem, “Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums,” *ACM Transactions on Information Systems*, vol. 26, 2008.
- [14] Theresa Wilson *et al.*, “OpinionFinder: A System for Subjectivity Analysis,” presented at the Human Language Technology Conference, 2005.
- [15] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment,” *The Fourth Annual International AAAI Conference on Weblogs and Social Media*, vol. 10, pp. 178–185, 2010.

- [16] R. P. Schumaker and H. Chen, “Predicting Stock Price Movement from Financial News Articles,” in *Information Systems for Global Financial Markets: Emerging Developments and Effects*, A. Yap, Ed. Hershey, PA: IGI Global, 2012, pp. 96–128.
- [17] K. Schouten and F. Frasincar, “Survey on Aspect-Level Sentiment Analysis,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2016.
- [18] A. Hogenboom, F. Frasincar, F. de Jong, and U. Kaymak, “Using Rhetorical Structure in Sentiment Analysis,” *Commun. ACM*, vol. 58, no. 7, pp. 69–77, Jun. 2015.
- [19] Anonymous, “Statistics and Facts on Sports Betting,” 2016. Retrieved July 13, 2016 from <http://www.statista.com/topics/1740/sports-betting>.
- [20] J. Culhane, “The DraftKings Crash,” 2015. Retrieved May 9, 2016 from http://www.slate.com/articles/sports/sports_nut/2015/10/the_insider_trading_scandals_cound_bring_down_draftkings_and_fanduel.html.
- [21] Jack Kerr, “The Dark Side of Sports Data,” 2016. Retrieved May 9, 2016 from <http://www.abc.net.au/news/2016-0107/kerr-the-darkside-of-sports-data/7072988>.
- [22] Dave Anderson, “Sports of the Times; Pro Football’s \$22 Roots,” *The New York Times*, 06-Jan-1983.
- [23] James Surowiecki, *The Wisdom of Crowds*. New York: Doubleday, 2004.
- [24] S. Luckner, J. Schroder, and C. Slamka, “On the Forecast Accuracy of Sports Prediction Markets,” in *Negotiation, Auctions, and Market Engineering*, vol. 2, H. Gimpel, N. Jennings, G. Kersten, A. Ockenfels, and C. Weinhardt, Eds. Berlin: Springer, 2008, pp. 227–234.
- [25] M. Spann and B. Skiera, “Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters,” *Journal of Forecasting*, vol. 28, pp. 55–72, 2008.
- [26] Daphne Raban and Sheizaf Rafaeli, “The Effect of Source Nature and Status on the Subjective Value of Information,” *Journal of the American Society for Information Science and Technology*, vol. 57, pp. 321–329, 2006.
- [27] Nancy Baldwin and Ronald Rice, “Information-seeking Behavior of Securities Analysts: Individuals and Institutional Influences, Information Sources and Channels, and Outcomes,” *Journal of the American Society for Information Science*, vol. 48, pp. 674–693, 1997.
- [28] R. Sullivan, A. Timmermann, and H. White, “Data-Snooping, Technical Trading Rule Performance, and the Bootstrap,” *The Journal of Finance*, vol. 54, pp. 1647–1691, 1999.
- [29] N. Jegadeesh, “Discussion,” *The Journal of Finance*, vol. 55, pp. 1765–1770, 2000.

- [30] L. Blume, D. Easley, and M. O'hara, "Market statistics and technical analysis: The role of volume," *The Journal of Finance*, vol. 49, pp. 153–181, 1994.
- [31] W.-C. Chiang, D. Enke, T. Wu, and R. Wang, "An adaptive stock index trading decision support system," *Expert Systems with Applications*, vol. 59, pp. 195–207, 2016.
- [32] A. Blair, "Guide to Charting," *Pitman, London*, 1996.
- [33] W.-K. Wong, M. Manzur, and B.-K. Chew, "How rewarding is technical analysis? Evidence from Singapore stock market," *Applied Financial Economics*, vol. 13, pp. 543–551, 2003.
- [34] W. Brock, J. Lakonishok, and B. LeBaron, "Simple technical trading rules and the stochastic properties of stock returns," *The Journal of Finance*, vol. 47, pp. 1731–1764, 1992.
- [35] R. P. Schumaker and H. Chen, "Evaluating a News-Aware Quantitative Trader: The Effects of Momentum and Contrarian Stock Selection Strategies," *Journal of the American Society for Information Science and Technology*, vol. 59, pp. 1–9, 2008.