PREDICTION BASED ON SENTIMENT ANALYSIS OF GAME

Jagdish Chandra Patni[1], Ravi Tomar[2], Mahendra Singh Aswal[3], Ankur Dumka[4]
[1,2,4]SoCSE UPES, Dehradun
[3]GKV, Haridwar

Abstract
The relationship between social media output and any game, using a dataset containing tweets from Twitter and game statistics. Specifically, we consider tweets pertaining to specific teams and games and use them alongside statistical game data to build predictive models of future game outcomes (which team will win?). Experiment with several features sets using large volumes of tweets and will try to match or exceed the performance of more traditional features that use game statistics. Our research motivation is to generate a system that can harness the wisdom of crowds using the sentiment information from Twitter to make match predictions

I. INTRODUCTION

We have always been involved in the thinking or opinions of "other" people for making the rightful decisions in our decision making processes, for e.g., for trying a new product, most people research upon it to view how many people agree or disagree with the quality, and then make decisions accordingly. Long before awareness of the World Wide Web became widespread, many of us asked our friends to recommend a scooter or to explain which team they are going to support, or consulted Consumer Reports to decide what vehicle to buy. But the World Wide Network has now (among other things) made it possible to find out about the opinions and neither experiences of those in the vast pool of people who are neither personal acquaintances nor well-known professional critics- common people of the world. Also, a huge number people are making their opinions available to strangers on social networking websites.

Sentiment analysis or opinion mining, is a subfield of computational linguistics concerned with extracting emotions from tweets, it also analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space.

Some interesting facts about the history of sentiment analysis are as follows:

- Sentiment Analysis grew out of Web integration Field.
- It started as an extension of knowledge extraction from data.
- This is the major reason for it being called as Opinion Mining.
- Many early algorithms mostly are extraction patterns.

As part of the effort of better organizing the information for users, researchers have been actively investigating the problem of automatic text categorization. A bulk of such work has been focused on topical categorization, attempting to sort documents according to their subject matter (e.g., sports vs. politics). However, recent years have seen rapid growth in on-line discussion groups, social networking websites such as Twitter, Facebook, Pinterest etc., and review sites (e.g., the New York Times' Books web page) where a crucial characteristic of the posted articles, is their sentiment, or overall opinion towards the subject matter — for example, whether a product review is positive or negative.

The term sentiment analysis perhaps first appeared in (Nasukawa and Yi, 2003), and the term opinion mining first appeared in (Dave, Lawrence and Pennock, 2003). However, the research on sentiments and opinions appeared earlier (Das and Chen, 2001; Morinaga et al., 2002; Pang, Lee and Vaithyanathan, 2002; Tong, 2001; Turney, 2002; Wiebe, 2000). In our project, we use the terms sentiment analysis and opinion mining interchangeably, we will use the term opinion to denote opinion, sentiment, evaluation, appraisal, attitude, and emotion. However, these concepts are not equivalent. The meaning of opinion itself is still very broad. Sentiment analysis and opinion mining mainly focuses on opinions which express or imply positive or negative sentiments. Linguistics and natural language processing (NLP) have a long history. Little research had been done about people's opinions and sentiments before the year 2000. Since then, the field has become a very active research area.

Although, linguistics and natural language processing (NLP) have a long history, little research had been done about people's opinions and sentiments before the year 2000. Since then, the sentiment analysis field has become a very active research area. There are several reasons for this. First, it has a wide range of applications, almost in every domain. The industry surrounding sentiment analysis has also flourished due to the proliferation of commercial applications. This provides a strong motivation for the project. Second, it offers many challenging interesting and intriguing problems, which had never been studied before.

The first launch of Twitter took place on July of 2006. Over the period of the last eight years, Twitter has become a big player in the social networking industry with an extremely large user-base, consisting of several millions of users. In particular, the number of Twitter accounts up to mid 2013, reached the number of 500 million user accounts. However the estimated number of active users was around 200 million, while the daily number of tweets, climbed up to 500 million. Twitter has been used for predicting or explaining a variety of other real-world events, such as the outcome of elections (Tumasjan, Sprenger, Sandner, & Welpe, 2010), the stock market (Bollen, Mao, & Zeng, 2011), box-office revenues of movies in advance of their release (Asur & Huberman, 2010) and the spread of diseases (Paul & Dredze, 2011). These are all strong evidence that Twitter can be a source of meaningful and useful information which can be exploited by using statistical methods.

- "Why aren't consumers buying our product?"
- We know the concrete data: price, specs, competition, etc.
- We want to know subjective data: "the design is tacky," "customer service was condescending"
- Misperceptions are also important, e.g. "updated drivers aren't available" (even though they are)

The essential issues in sentiment analysis are to identify how sentiments are expressed in tweets and whether the expressions indicate positive (favorable) or negative (unfavorable) opinions toward the subject. In order to improve the accuracy of the sentiment analysis, it is quite important to properly identify the semantic relationships between the sentiment expressions and the subject. By applying semantic analysis with a syntactic parser and sentiment lexicon, our prototype system achieved high precision in finding sentiments within the tweets.

An end user might desire an automated at-a-glance presentation of the main points made in a single review or how opinion changes time to time over a time period of the tournaments. To meet the requirement we will use the 5W task which seeks to extract the semantic constituents in a natural language sentence by distilling it into the answers to the 5W questions: Who, What, When, Where and Why. The visualization system facilitates users to generate sentiment tracking with textual summary and sentiment polarity wise graph based on any dimension or combination of dimensions as they want i.e. "Who" are the actors and "What" are their sentiment regarding any topic, changes in sentiment during -"When" and "Where" and the reasons for change in sentiment - "Why".

As we know that necessity is the mother of all Invention and in today's digital age, text is the primary medium of representing and communicating information, as evidenced by the pervasiveness of e-mails, instant messages, documents, weblogs, news articles, homepages and printed materials. Our lives are now saturated with textual information and opionions, and there is an increasing urgency to develop technology to help us manage and make sense of the resulting information overload. While expert systems have enjoyed some success in assisting information retrieval, data mining, and natural language processing (NLP) systems, there is a growing requirement of Sentiment Analysis (SA) system which can process automatically the plethora of sentimental information available in online electronic text. So far the Sentiment Analysis research becomes quite mature after a few decade of cultivation. The focus of this project is on aggregating and representing sentiment information drawn from an individual tweet or from a collection of tweets. Sentiment/opinion aggregation is necessary requirement at the end user's perspective, for example, an end user might desire an at a glance presentation of the main points made in a single review or how opinion changes time to time over multiple matches. On real-life applications, to provide a completely automated solution is the ultimate desired goal of all the sentiment analysis research. An intelligent system should smart enough to aggregate all the scattered sentimental information from the various tweets, which are based upon news article and written reviews. The role of any automatic system is to minimize human user's effort and produce a good sensible output. There is no doubt that aggregation of sentiment is necessary but it is very hard to find out the common consensus among researchers that how the sentimental information should be aggregated. Although a few systems like Twitter Sentiment Analysis Tool1, TweetFeel2 are available in World Wide Web (WWW) since last few years still more research efforts are necessary to match the user satisfaction level and the social need.

Tweets and texts are short mainly including a sentence or a headline rather than a document. The language used is very informal, with creative spelling and punctuation, misspellings, slang, new words, URLs, and genre-specific terminology and abbreviations, such as, RT for "re-tweet" and # hashtags, which are a type of tagging for Twitter messages. How to handle such challenges so as to automatically mine and understand the opinions and sentiments that people are communicating has only very recently been the subject of research for e.g.,Jansen et al., 2009 ; Barbosa and Feng, 2010;

Bifet and Frank, 2010; Davidov et al., 2010; OâĂŹConnor et al., 2010; Pak and Paroubek, 2010; Tumasjen et al., 2010; Kouloumpis et al., 2011). Another aspect of social media data such as Twitter messages is that it includes rich structured information about the individuals involved in the communication. Twitter mainly maintains information of who follows whom and re-tweets and tags inside of tweets provide discourse information. Modelling such structured information is important because: it can lead to more accurate tools for extracting semantic information, and also because it provides means for empirically studying properties of social interactions (e.g., we can study properties of persuasive language or what properties are associated with influential users).

2. Literature Survey

Some research has already been done in the areas of Sentiment analysis applications and traffic safety. The following sections will give a brief overview of some of this research, which will aid in the research for this project.

2.1. The Cricket Tweet

The project focuses on applying NLP techniques to do analysis of relevant, unstructured data of tweets on a given match day of cricket. The game itself makes the project interesting. There were around 64 million tweets in the last week about the game and around 101.7 million people watched the game and many of them expressed their opinions and emotions about the game on Twitter via tweets. Twitter also has official hash tags for IPL (#IPL, #IPL2015) and attracts a lot of people all around the world because of the extensive popularity of the game. Our solution can be used by marketing companies to increment user engagement on the targeted website. Also, the researchers of Human Behavior can make a difference by analyzing the pattern changes in the human reactions during the various events of the game. The business analysts can monetize the trends to increase the website traffic. The International Cricket Council (ICC) can broadcast the results on television and websites to engage more users and hence the income. The major challenge faced by us includes annotating the tweets because of their unstructured and semi-formal nature and because of our different angles to look at a tweet as two different people can annotate a particular tweet differently based on the individual nature. We had to do a lot of manual analysis in order to get the results evaluated. Evaluation of models was difficult as every match presented different set of data.

2.2 Predicting the NFL Using Twitter

In this paper they studied the relationship between social media and National Football League (NFL) games, using the data containing tweets from Twitter and NFL game statistics. Specifically, they considered tweets relating to specific teams and games in the NFL season and use them together with statistical game data to build prognostic models for future game outcomes that iswhich team will win and sports betting outcomes that is basically which team will win with the point spread? will the total points be over/under the line?). They experimented with numerous feature sets and find out that simple features using large volumes of tweets can match or exceed the performance of more traditional features that use game statistics.

2.3 Sentiment Analysis of Twitter on English Premier League Soccer

In this paper they studied the twitter feeds of the EPL and found out that twitter to be a better predictor of the game outcome than the experts and also the overall ranking. The result was theoretically grounded that crowd wisdom can be a better predictor than individuals. They presented the CentralSport system that use the sentiment analysis of tweets to predict the EPL football outcomes.

2.4. Tweepy Documentation

Tweepy Documentation helps to understand the APIs provided for collecting Twitter data. The Advantages of Careful Seeding explains the benefits of using k-means++ over k-means as the prior one takes into account pre-defined cluster centroids. The paper "Summarizing Sporting Events Using Twitter" specifies methods for tweet summarization which takes into account chunking of the tweets based on time stamp and summarizing each chunk. On top of that we added our average function to average out all the tweets from the chunks to find a threshold value above which a chunk is considered as a peak valued chunk. Only from these chunks we summarized top scored tweets to give better summary.

2.5. Stock Prediction Using Twitter Sentiment Analysis

In this research paper sentiment analysis and machine rules were applied in order to make a relationship between Public and Market Sentiment. Twitter data was used to forecast public mood and use the expected mood and preceding days' DJIA values were used to predict the stock market activities. For analysing our outcomes, a new cross authentication method was proposed for financial figures and 75.56% precision was obtained using Self Organizing Fuzzy Neural Networks (SOFNN) on the Twitter feeds. Another simple portfolio management strategy was executed grounded on the anticipated values. Our work is grounded on Bollen et al's well-known paper which anticipated the same with 87% precision.

2.6. Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement

This research paper investigated the association between Twitter feed content and stock market programme. Specifically, there is a hope to see if, and how sound, sentiment evidence pull out from these feeds can be used to forecast forthcoming alterations in prices. In order to respond to this query, we create a prototype, guesstimate its accurateness, and put it to the trial on existent market statistics by means of a simulated portfolio. Our outcomes specify that the prototype is effective in spawning further profit.

2.7. News Analytics and Sentiment Analysis to Predict Stock Price Trends

Business news transmits diverse data of diverse corporations. But in this promptly moving world the amount of news foundations existing is innumerable, and it's feasibly impossible to deliver and discover all appropriate material in the form of news to pull a assumption apt to make an investment strategy that yields supreme revenue. This research paper proposed a analytical prototype to forecast sentiment about stock price. First the applicable real time bulletin headings and press-issues have been strained from the huge set of business news foundations, and then they have been analyzed to forecast the sentiment about corporations. In order to find association between sentiment projected from news and original stock price and to check effective market postulate, we design the sentimentalities of 15 odd firms over a period of 1 month. Our outcome displays an normal precision score for recognizing precise sentiment of nearby 70.1%. We also have strategized the faults of forecast for diverse corporations which have got out the RMSE and MAE of 30.3% and 30.04% respectively and an improved F1 factor of 78.1%. The assessment between positive sentiment curve and stock price trends discloses 67% co-relation between them, which specifies to presence of a semi-strong to strong well-organized market assumption.

2.8. Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation

Though uninteresting individually, Twitter messages, or tweets, can provide an accurate reflection of public sentiment on when taken in aggregation. In this paper, we primarily examine the effectiveness of various machine learning techniques on providing a positive or negative sentiment on a tweet corpus. Additionally, we apply extracted twitter sentiment to accomplish two tasks. We first look for a correlation between twitter sentiment and stock prices. Secondly, we determine which words in tweets correlate to changes in stock prices by doing a post analysis of price change and tweets. We accomplish this by mining tweets using Twitter's search API and subsequently processing them for analysis. For the task of determining sentiment, we test the effectiveness of three machine learning techniques: Naive Bayes classification, Maximum Entropy classification, and Support Vector Machines. We discover that SVMs give the highest consistent accuracy through cross validation, but not by much. Additionally, we discuss various approaches in training these classifiers. We then apply our findings to on an intra-day market scale to find that there is very little direct correlation between stock prices and tweet sentiment on specifically an intra-day scale. Next, we improve on the keyword search approach by reverse correlating stock prices to individual words in tweets, finding, reasonably, that certain keywords are more correlated with changes in stock prices. Lastly, we discuss various challenges posed by looking at twitter for performing stock predictions.

2.9. Sentiment analysis of news articles for financial signal prediction

Due to the instability of the stock market, price fluctuations grounded on sentiment and bulletin stories are common. Dealers attract upon a varied range of openly-accessible data to notify their market conclusions. For this assignment, we dedicated on the examination of publicly-available news reports with the use of computers to deliver guidance to brokers for stock transaction. We established Java data handling code and used the Stanford Classifier to rapidly examine financial bulletin courses from The New York Times and forecast sentiment in the courses. Two methodologies were taken to yield sentiments for training and testing. A labour-intensive methodology was tried by means of a human to deliver the articles and categorising the sentiment, and the automatic tactic using market activities was used. These sentiments could be used to forecast the day-to-day market tendency of the Standard & Poor's 500 index or as inputs to a bigger transaction system.

3. Model and Architecture

There has been a lot of work done on Tweet classification into sentiments, but those works mostly include positive, negative and neutral sentiments. We came across several papers that discuss about summarization of the tweets.We modified the concept a bit by including our own concepts/ideas and boosting the tweet scores further up.

Tweepy Documentation, which is also a package used by our python code helps to understand the APIs provided for collecting Twitter data. The Advantages of Careful Seeding explains the benefits of using k-means++ over k-means as the previous one takes into account pre-defined cluster centroids. The paper "Summarizing Sporting Events Using Twitter" tells us about the methods for tweet summarization which takes into account the division of the chunks of the tweets based on time stamp and summarizing each chunk. In addition to that we added our average function to average out all the tweets from the chunks to find a threshold value above which a chunk will be considered as a peak valued chunk. Now from these chunks we summarized top scored tweets to give better summary.The base origin of data (here tweets) is twitter. Tweets had to be captured from twitter at real time when the match began. Option was to use Streaming API of twitter and ingest the tweets. We used a wrapper project of this Streaming API (Tweepy) to build our system to capture tweets on real time basis. Other data set required was for the gazettes used.The data for NER (Players, Location, Teams, Venue, Stadium) were scrapped manually from the official website of Indian Premiere League 2015. Further known events were formed a part of gazettes and domain experts (here self) made available events for the game of Cricket for IPL game format. Size and other details: Tweets of 3 different matches were captured in 3 different categories(Match_1, Match_2, Match_3). The size of each category called as bucket is around

8,000 to 10,000 tweets. The tweets were only filtered for English language using the API configuration. They were further sub-divided in data store containing maximum of 200 tweets each just to be safe if any debugging would be required on raw data in future. Every tweet has all the attributes like tweet Id, time stamp, tweet text etc available in JSON format in the data set.

3.1 DATA COLLECTION AND PREPROCESSING

Size of data and other details:

Tweets of 3 different matches were captured in 3 different categories buckets (Match_1, Match_2, Match_3). The size of each category called as bucket is around 8,000 to 10,000 tweets. The tweets were only filtered for English language using the API configuration. These buckets are further sub-divided in data store containing maximum of 200 tweets each just to be safe in case if any debugging would be required on raw data in future. Each tweet has all the attributes like tweet Id, time stamp, tweet text etc available in JSON format in the data set.

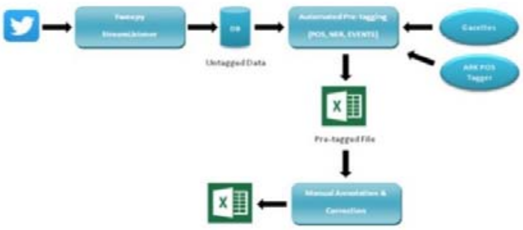

Fig 3.1 ***Data Collection***

3.2. SENTIMENT ANALYSIS

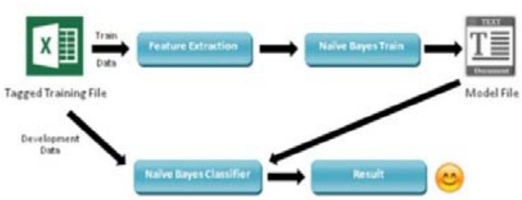

Fig 3.2The Flow Diagram of sentiment analysis

We have defined our own sentiment classes as Unpleasant, Sad, Neutral, Happy, Ecstatic. Before forming a training set, we took out some of the words which were proper nouns, determiners which really don't contribute to the classification task. As we have followed Naive Bayes approach, the keywords are added in the bag of words feature set and the model is trained on that data. We had round about 1000 tweets given as a training dataset. A development data is then provided along with the model file to the classifier to get sentiment classified tweets. The results are then post processed to generate a JSON to be given to the Data Driven Document object in order to get the graphical representation of the results. We have tested our results using four approaches - Megam Multiclass Classifier, Naive Bayes Classifier, NLTK Naive Bayes Classifier and Bi-gram Naive Bayes Classifier and hence have formed the evaluation matrix.

3.3 Recognition of Named Entity

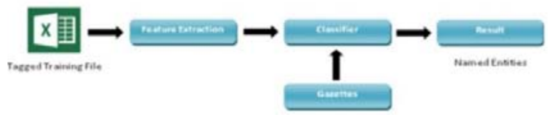

Fig 3.3 Recognition of Named Entity Flow

The features used for Named Entity Recognition (NER) task are fetched from the pre-tagged training file for Named Entities. We have defined our own named entities as Person, Location, Team and Venue. The tagged named entities from the training file are given as input to the classification module and on top of it; we have defined Gazettes for all of the four named entities which consist of all the players, locations, teams and venue names along with the acronyms and pet names of the players. The classification module then gives the separate named entities for all the four classes. The output is processed to form a JSON file which is then given as an input to Data Driven Document module for graphical representation.

3.4 Clustering

The aim of clustering is to explore something interesting in an unexplored data set. Further re the aim has been refined to cluster using NLP techniques on known set of events that were previously tagged. In the k-means clustering with the help of initial seeds to known event tweets, we have built our own k-means clustering algorithm based on centroid model clustering and initialization using known k-events.

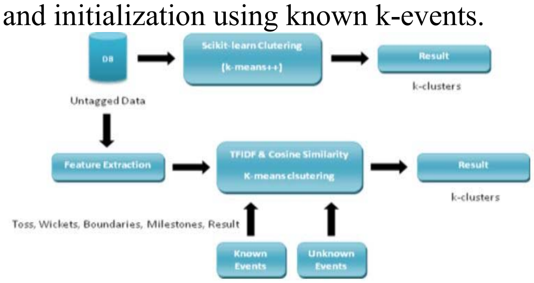

Fig 3.4 ***k-mean++ Clustering***

3.5 Initialization using the seed tweets :

Tagged the tweets with different events (TOSS, BOUNDARY, WICKET, RESULT, and MILESTONE). We have randomly selected 1 tweet from each event and form an initialization cluster with centroids mapping to the vector representation of the tf-idf values of tokens filtered. Features are extracted to remove stop words, and few tokens matching POS tags (!, ', #, @, P). This feature representation forms a bag of words and we have created a tf-idf vector representation for each tweet and have applied the k-means algorithm with the above initialization setup. For similarity we have also calculated the cosine distance between the tweets and matched tweets to nearest clusters.

The convergence of the algorithm will take place at around at 6th iteration for around 8000 tweets and k=5 (number of unique events). Off-the-shelf technique has been used to explore random clusters: We have also used scikit learn k-means++ module to visualize the data for clusters. The clusters start at unknown seeds, so it's very difficult to predict the evenness of the tweets in one cluster. There were few interesting clusters explored which formed clusters on some named entities, some were on events too while few were random with some similar features.

3.5 Summarization

Summarize an event based on social context available. But very few have tried to explore the game of cricket. We referred a paper "Summarizing Sporting Events Using Twitter" by Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. This approach is based on finding the peaks in the frequency of tweets and applying the summarization technique only on the tweets in the window of peaks. Finally merging all the intermediate summaries to form a match summary. Initially we read the raw dump of tweets as explained in the data collection, and formed timeline model of tweets. The tweets in our case were bucketed in a bucket size of "m"-minutes. The number of tweets in each "m"-minute bucket has some number of tweets which can vary from bucket to bucket. We have selected the value m = 3 min bucket. Thus data gets chunked on timeline series with 3- minutes bucket. Next task is to identify the peaks in the time lined buckets, this was found based on average frequency of tweets as threshold and mark only those buckets that have number of tweets more than this threshold. The resultant is an array of consecutive-chunks each consecutive chunk depicting something exciting happening. The tweets from each consecutive chunk are considered as a data-set for clustering on known events (TOSS, BOUNDARY, WICKET, RESULT, and MILESTONE). Now finding which event has really occurred was done using k-means with initial seeds as above method explained and then applying scores to those tweets that are more relevant to the maximum frequency words (add +1 since it has some relevancy to the cluster) and the known events (add +5 to score since more specific to an event).

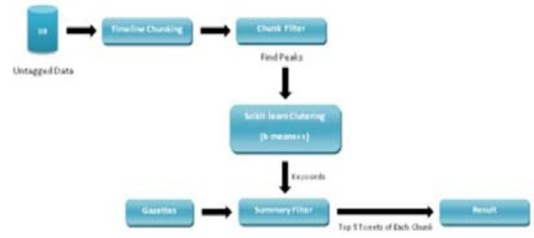

Fig 3.5 Summarization Flow

Then rank the tweets based on the final score and take the top-5 tweets to summarize the event happening in the consecutive-chunk selected. This was repeated to all the consecutive chunks in the array as mentioned above. Finally all the results are combined to generalize the overall summary of the match. We found the observations interesting for finding events like TOSS at the start, wickets, and mile-stones during the course of match and then a result towards the end with the man-of-match.

3.6 Evaluation and Analysis

In the sentiment Analysis, initially we had started with 9 sentiment classes and ended up with accuracy around 60%. So, we decided to combine or merge the classes and finalized 5 sentiment classes. We have followed an incremental approach by training the dataset on a chunk of 200 tweets at a time and classifying the tweets to identify trends in the accuracy level. Our baseline for sentiment analysis resulted from the Naive Bayes classification for the first assignment. Although, the data here was unstructured, we did comparatively well with the accuracy. We wrote the accuracy calculation script, and also tested the accuracy on development dataset.

In the Named Entity Recognition (NER), we evaluated the classified named entities by crosschecking them with our predefined gazettes for Person, Locations, Teams and Venue. After collecting data from the official IPL site, the files were formed. The major challenge while evaluating named entities was to consider short names and acronyms as well. We did fairly enough to consider all these scenarios while forming gazettes.

In the Clustering, evaluation of Clusters formed was more of an exploratory task. We went through several clusters and categorized our results based on the clustered tweets similarity. We also cross checked the relevancy of result with our pre-defined event based clusters. Sometimes, the clustered data is not relevant to our interest since it is purely based on text similarity. In the summarization, we evaluated the summaries generated was based on the manual analysis and comparing the summaries with the online sport featuring channels like cricbuzz.com. Summarization is based on similarity between the tweets and their scores were based on the events. So, we did get some irrelevant summaries from the clusters. The summaries are given as a set of five highly relevant tweets. Our future work now involves summarizing these tweets into a one line summary using Parse Graphs.

3.7Implementations

In this section we are discussing how we are going to implement our idea into reality. For the project, we suggested a model for the logistic reversion classifier to predict games. In order to measure the performance of our feature sets, and tune hyperparameters for our model as the season progresses. For our calculation we used the Indian Premier League season. Match results were obtained from espncricinfo.com. Tweets were extracted from the Twitter streaming API using team-specific hashtags. Each team had one hashtag recognized from http://coolestguidesontheplanet. com/indianpremierleague-twitter-hashtags. For our study period, 18,454,362 tweets were gathered. The average tweet length was 106.0 characters. Table 1 shows the analysis of the number of tweets by team.

From this table, Chennai Super Kings, Kolkata Knight Riders, Royal Challengers Bangalore, had the most tweets. For each match we used tweets from the match days right up to match start. From these data we constructed one distinct model that is sentiment. We further constructed eight models investigating the roles of tone and polarity.

Team Name	#Tweets
Chennai Super Kings	5,459,577
Rajasthan Royals	3,063,851
Royal Challengers Bangalore	3,969,053
Kolkata Knight Riders	5,748,831
Mumbai Indians	2,300,352
Delhi Daredevils	1,037,740
Kings XI Punjab	1,138,901
Sunrisers Hyderabad	1,256,165

Table 1. Breakdown of the number of tweets and tweeters by team

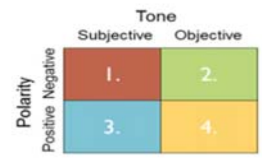

Table 2. Sentiment Models

In the following table have total eight models:
Model 1 – Subjective Negative tweets,
Model 2 – 0bjective Negative,
Model 3 – Subjective Positive,
Model 4 – 0bjective Positive,
Model 5 – All Subjective,
Model 6 – All 0bjective,
Model 7 – All Negative, and
Model 8 – All Positive.

Tone evaluates whether a tweet is subjective or objective. It is a binary classification made by 0pinionFinder on each sentence. For tweets with more than one sentence, we use majority rules. In cases of a tie or ambiguity the sentence is marked neutral and we don’t consider it further.

Model	Accuracy
Model 1	65.57%
Model 2	64.86%
Model 3	52.00%
Model 4	66.67%
Model 5	52.54%
Model 6	48.65%
Model 7	43.44%
Model 8	50.58%

Table 6.1: Accuracy results of models

4. Conclusion

Sentiment is a predictor of match and tournament outcomes. In terms of accuracy all sentiment models beat random chance and Model 3 – Subjective Positive outperformed the odds-only Baseline model (66.67% versus 65.57% respectively). Digging further into Model 3’s predictions we found that prediction accuracy for

Home and Away outcomes averaged 76.8% and this model was hurt by Draw outcomes, for which it did not 15 account. For wagering payout, six of the eight sentiment models made an excess return and three models outperformed odds-only Baseline. We speculate that polarity influenced outcomes and that the addition of further tonal or polar sentiment may have harmed the results of the other models. Clearly more research could be done in this area. There are many potential extensions to this research as the system we created could be ported to other sports domains. One such extension would be the inclusion of Draw categories. While we ignored this category in our study and still managed good results, future work should look into ways of algorithmically identifying Draws with good accuracy. Another extension would be to analyze tweets during a match or briefly thereafter. It might provide some additional insight into tweeter behavior based on goal differences such as more/less interest in matches with more/less goal differentials. A third extension would be to analyze tweets and performance in the first half of the season versus the second half. Sinha et. al. discovered season-half differences in the IPL and perhaps similar differences exist in the Premiership. Fourth, an analysis of tweets and retweets may prove interesting in answering the question of what conditions lead to the most retweets? Lastly, it would be interesting to merge sentiment and social network theory to identify the tweeters with the greatest say on setting the sentiment mood for a particular hashtag. It is quite clear that within this domain there are plenty of opportunities for further research.

References:

[1] J. Surowiecki, The wisdom of crowds: Doubleday, 2004 .

[2] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, pp. 1-8, 2011.

[3] Tumasjan, Andranik, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." ICWSM 10 (2010): 178-185.

[4] Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on, vol. 1, pp. 492-499. IEEE, 2010.

[5] Bothos, Efthimios, Dimitris Apostolou, and Gregoris Mentzas. "Using social media to predict future events with agent-based markets." (2010).

[6] Sinha, Shiladitya, Chris Dyer, Kevin Gimpel, and Noah A. Smith. "Predicting the NFL using Twitter." arXiv preprint arXiv:1310.6998 (2013).

[7] Hong, Yancheng, and Steven Skiena. "The wisdom of bookies? sentiment analysis vs. the NFL point spread." In Proceedings of the international conference on Weblogs and Social media (icWSm-2010). 2010.

[8] Hotz, R. L. "Decoding our chatter." TECHNOLOGY (2011).

[9] R. P. Schumaker, O. K. Solieman, and H. Chen, "Sports knowledge management and data mining," Annual review of information science and technology, vol. 44, pp. 115-157, 2010.

[10] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, pp. 1-8, 2011.