

Predicting Indian Premier League (IPL) Cricket Match Results Using Social Media Analytics : An Opinion Mining Approach

Sanjeev Chauhan¹, Dr. Gunjan Mohan Sharma², Dr. Vikas Tripathi³

¹Assistant Professor, IBM, GLA University, sanjeev.chauhan@gla.ac.in

²Associate Professor, Jindal Global Business School, OP Jindal Global University,
gmsharma@jgu.edu.in

³Professor, IBM, GLA University, vikas.tripathi@gla.ac.in

Abstract

With the advent of Web 2.0, technology has taken a leap and penetrated into every human being's life and people tend to share in a social space, in fact it has become the prominent choice of expression of feelings for the millennials. The most popular discussion topic would be sports and entertainment as the trend has taken a paradigm shift with the launch of online streaming platforms. In this article, I am trying to identify what do people discuss about Indian Premier League Cricket and its teams in an online space and trying to figure out the feasibility in using the sentiments or opinions from posts made in various social media platforms to predict the winner of a Match and eventually the League, thereby an effort to create a predictive mechanism to support the various betting and fantasy games like Dream 11. For predicting the results of a match, one can identify the number of posts made before a match for each of the team and the sentiments associated with the posts made in Facebook, blogs and other public forums. We can use this method to identify the winning team before the start of the match and the efficiency and precision of the method is calculated using the Support Vector Machine (SVM) which has been used in various seminal papers as the preferred classifier.

Keywords: Sentiment Analysis, Indian Premier League, IPL Predictor, Opinion Mining

1. INTRODUCTION:

It's a known fact that Betting is illegal. Though Indian Cricket players haven't been involved in the gambling issues like Pakistani and South-African cricketers have. India is the market being targeted by the gamblers due to the huge influx of money churned in the same (Mahyera, 2012). This is to be curbed out by the government through legalising the betting process and tracking the sources there in. However, we saw the acceptance of legal betting websites such as Dream 11 (currently 20 million customer base) where people happily bet and choose their best XI to win cash. There isn't any mechanism to help the participants know which team is having higher chance to win as the winning side team members will be getting a higher bonus. Apart from these, there are websites where betting happens solely on the basis of predicting the winner of the match (Ferrucci, 2012). The research was conducted during April- May, 2018. The research problem is to Identify which team is to bet for before an Indian Premier League match based on the positive sentiments present in the social media posts by fans. This could not only serve as a mechanism for the Cricket matches as a successful model could help one in reaping benefits out of the number of emerging leagues in the Indian Sports context such as Indian Super League, Kabaddi League, Badminton League, Etc. The higher precedence in the minds of people would fetch a larger dataset to predict the winners of the matches. This could also be used by Companies and Brands while endorsing a particular team before the start of a tournament to get a higher penetration into that Team's fans and followers (Nyer, P. U., 2000). The key factor here is the amount of posts being made by the fans in an online space through their social media profiles, this could be generated more with the customer engagement activities by the team franchisee itself or by the tournament organisers themselves. A higher trend line and mentions would help the tournament to fetch higher funds through advertisements and endorsements.

2. METHODOLOGY

The tool used for the analysis is IBM Watson Social Media Analytics by which the sentiments associated with each team in an upcoming match was identified to predict the result of the match. Watson Social Media Analytics makes it possible to choose a time-frame within which the data is to be extracted from (Ferrucci, 2012). This was essential to identify the social media buzz about a particular team before a said match until the time of the play starts (in IPL this being 3:59 pm IST on weekends and 7:59 pm IST on weekdays). Watson would also mention how much percentage of mentions of a topic is positive and negative. This is utilized by using the team names as Topics to find the sentiments and using the City name associated with each team as a Theme to identify the mentions. Also, the tool <https://netlytic.org/> was used to collect tweets on hashtags of various teams, but it comes with a limitation of 1000 latest tweets, which makes it not usable for the whole study. The regional languages would be converted to ASCII codes which makes it difficult to clean the data in Excel. Hence, not used for the part of the report.

3. DATA CREATION:

For the purpose of this report, a total of 13 Projects were created in IBM Watson Social and Sentiment analytics. The naming convention for the Project has been “IPL - <Match-Date>”. Out of the 52 matches that has occurred till date on 19th May, 2018 20 of them has been chosen randomly as a sample to test the validity of predicting the match results through Sentiment Analysis. Following are the Dates chosen for the report:

- 14th, 15th, 21st, 22nd and 28th April 2018
- 5th, 6th, 10th, 12th -15th May 2018

In each project, the topic would include: Indian Premier League and the respective teams having the Match on that particular day from the following list:

- Chennai Super Kings (CSK), Delhi Daredevils (DD), Kings XI Punjab (KXIP), Kolkata Knight Riders (KKR), Mumbai Indians (MI), Rajasthan Royals (RR), Royal Challengers Bangalore (RCB) and Sunrisers Hyderabad (SRH).

On projects where only a single match is played, the topics would include <Team1> vs <Team2>. For example, the match between Rajasthan Royals and Kolkata Knight Riders would have topic RRvsKKR and KKRvsRR, in order to find more documents.

Reliability: Following the Topics, Themes are chosen as the City names associated with each of the franchisees. Namely, Chennai, Rajasthan, Delhi, Punjab, Kolkata, Mumbai, Bangalore and Hyderabad. Mentions with no themes averaged between 30 % to 40 %.

Recency: The time period chosen for each of the project in One day. Starting from 1 am IST of the day previous to the match day till either 3:59 pm or 7:59 pm IST before the commencement of the match to identify the recent sentiment just before start of the play.

4. DATA ANALYSIS

The sentiment of the Data collected was checked to identify the degree of positive sentiments associated with each team. However, after the initial analysis, the sentiments were edited to correct the various discrepancies in IBM Watson to identify certain wrong sentiments marked, such as “attacked”, “batting”, etc as these were marked under negative sentiments. The following table of words were used as reference for editing sentiments (taken from Mustafa et al. 2017):

Positive	Negative
win, victory, cruise, top, good, strength, brilliant, phenomenal, wow, patchy, well done, excellent, nice, majestic, terrific, deserved, spectacular, high-class, awesome, wonderful, best, delight, comeback, massive, perfect, favorite, outstanding, great, dominate, flatten, ovation, convincing, blinder, emphatic, magic, promising, flawless, classic, likely, influential, all the best, strong, celebration, impressive, well played, cheer, out played, splendid, mighty, superb, nailed, triumph, fabulous, astounding, solid, sufficient, good luck, encouraging, matchless, sublime.	loss, defeat, suck, bad, crumble, terrible, unwatchable, awful, boring, stupid, worse, waste, poor, torment, unconvincing, hopeless, regret, thrash, disappointing, furious, alas, vulnerable, blow, worthless, unlikely, destroy, pointless, hard luck, unfortunate, dismantle, wreck, deprive, ruin, dismal, over, needless, mockery, unlikely, weak, all over, disaster, shabby, amateur, unimpressive, wasted, collapse, bad luck, flop, unexciting, typical, falling short, unachievable, unable, exposed, setback, unbelievable, trouble, crush, shaky.

The analysis of various sentiments on a project-to-project are mentioned below. Post the analysis, a Support Vector Machine (SVM) proportion is used to predict the winner of the tournament (Mustafa et al. 2017) based on the f-measure calculated based on the precision of predictions till date and recall of the past performance. As per the model suggested by Mustafa et al. the following formulae are adopted for evaluating the values of prediction:

$$recall = \frac{\text{number of instances that are predicted as positive}}{\text{number of positive instances}} \quad (10)$$

$$precision = \frac{\text{number of instances that are predicted as positive}}{\text{total number of positive predictions}} \quad (11)$$

$$f\text{-measure} = \frac{2 \times precision \times recall}{precision + recall} \quad (12)$$

5. MATCH WISE ANALYSIS THROUGH PROJECTS:

- Match Date: 14/04/2018
 Matches:
 (1) KKR vs SRH
 (2) MI vs DD
 Prediction based on Model: (1) SRH (2) DD
 Actual Winners: (1) SRH (2) DD

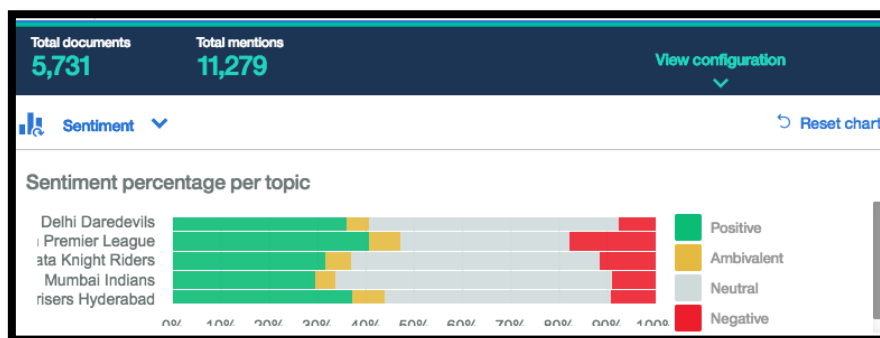


Figure 1. Social mention insights for Match Date: 14/04/2018

2. **Match Date:** 15/04/2018

Matches:

- (1) RCB vs RR
- (2) KXIP vs CSK

Prediction based on Model: (1) RR (2) KXIP

Actual Winners: (1) RR (2) KXIP

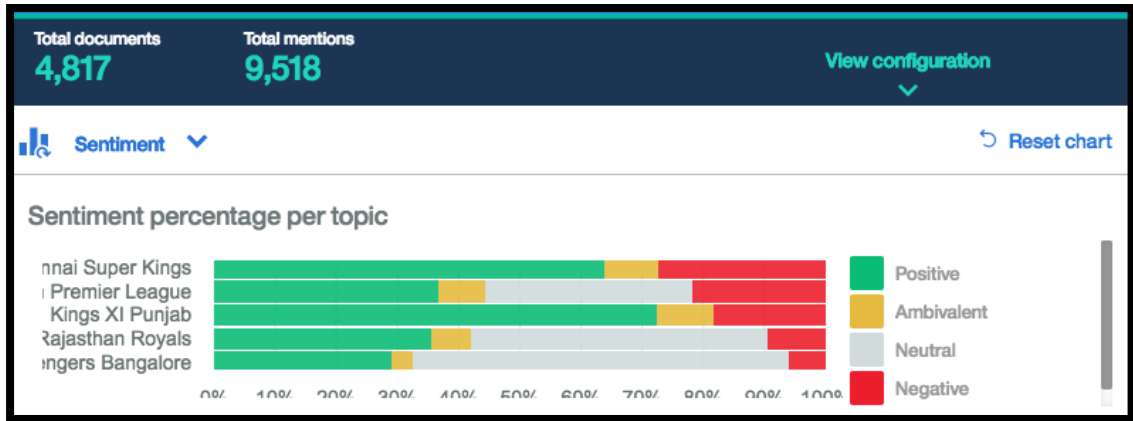


Figure 2. Social mention insights for Match Date: 15/04/2018

3. **Match Date:** 21/04/2018

Matches:

- (1) RCB vs DD
- (2) KKR vs KXIP

Prediction based on Model: (1) DD (2) KXIP

Actual Winners: (1) RCB (2) KXIP

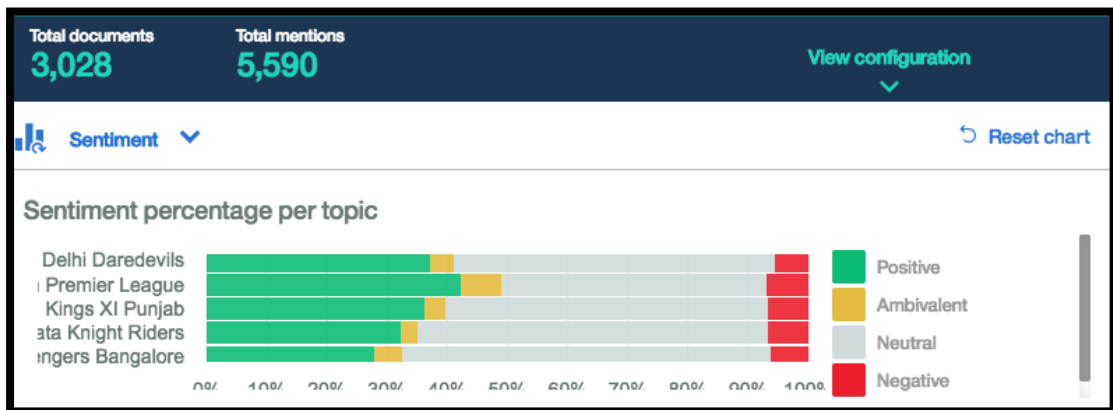


Figure 3. Social mention insights for Match Date: 21/04/2018

4. **Match Date:** 22/04/2018

Matches:

- (1) SRH vs CSK
- (2) RR vs MI

Prediction based on Model: (1) SRH (2) RR

Actual Winners: (1) CSK (2) RR

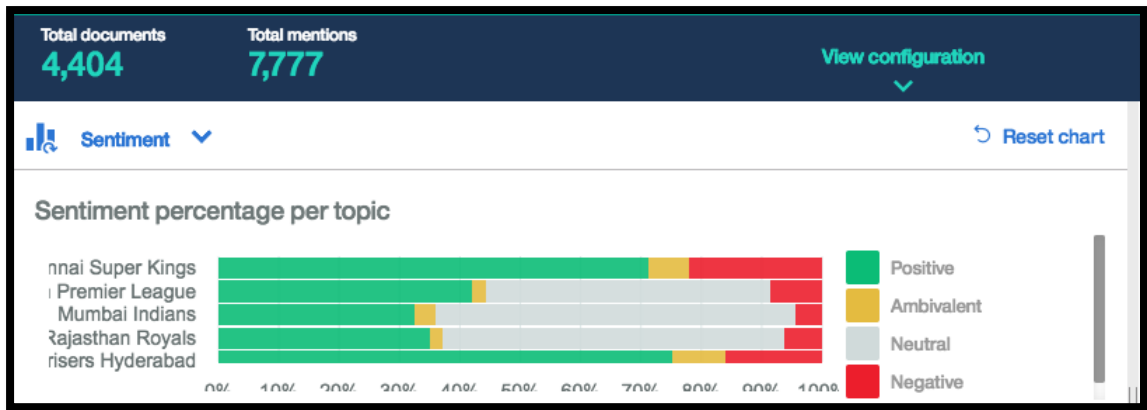


Figure 4. Social mention insights for Match Date: 22/04/2018

- Match Date:** 28/04/2018
Match:
 (1) MI vs CSK
Prediction based on Model: (1) MI
Actual Winner: (1) MI

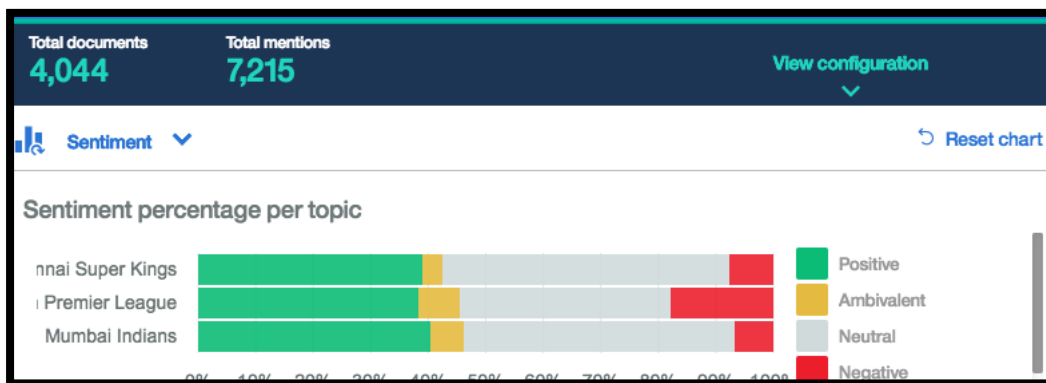


Figure 5. Social mention insights for Match Date: 28/04/2018

- Match Date:** 05/05/2018
Matches:
 (1) SRH vs DD
 (2) CSK vs RCB
Prediction based on Model: (1) SRH (2) CSK
Actual Winners: (1) SRH (2) CSK



Figure 6. Social mention insights for Match Date: 05/05/2018

7. **Match Date:** 06/05/2018

Matches:

- (1) MI vs KKR
- (2) KXIP vs RR

Prediction based on Model: (1) KKR (2) KXIP

Actual Winners: (1) MI (2) KXIP

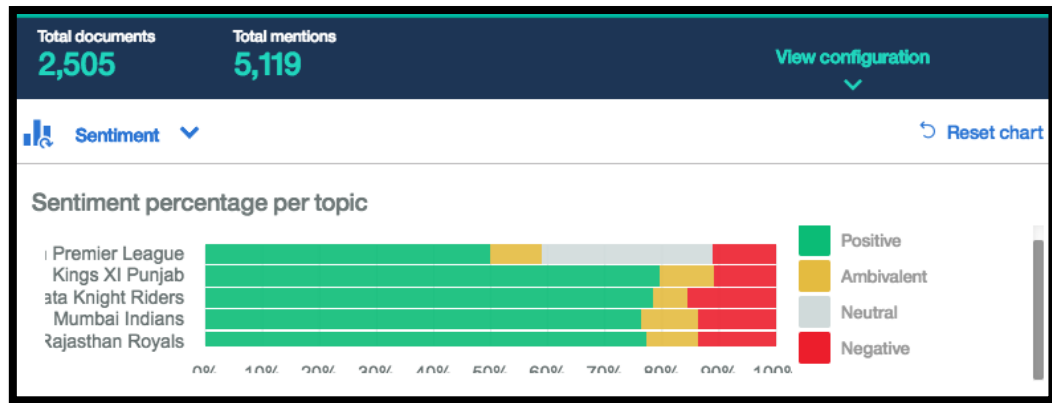


Figure 7. Social mention insights for Match Date: 06/05/2018

8. **Match Date:** 10/05/2018

Match:

- (1) DD vs SRH

Prediction based on Model: (1) SRH

Actual Winner: (1) SRH

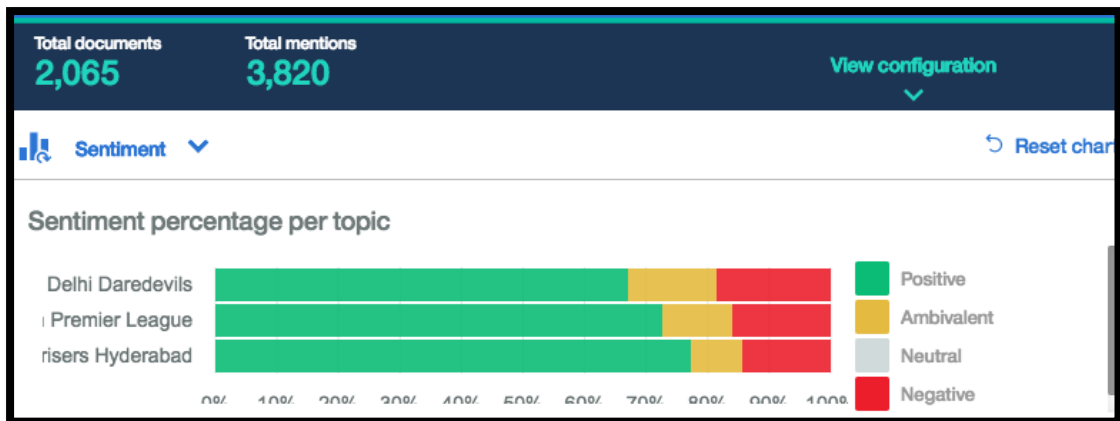


Figure 8. Social mention insights for Match Date: 10/05/2018

9. **Match Date:** 12/05/2018

Matches:

- (1) DD vs RCB
- (2) KXIP vs KKR

Prediction based on Model: (1) DD (2) KKR

Actual Winners: (1) RCB (2) KKR

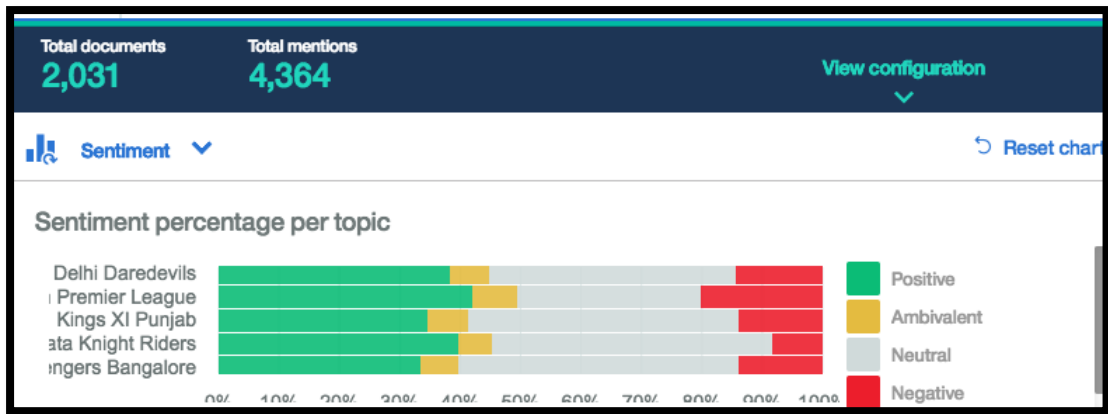


Figure 9. Social mention insights for Match Date: 12/05/2018

10. **Match Date:** 13/05/2018

Matches:

- (1) MI vs RR
- (2) CSK vs SRH

Prediction based on Model: (1) RR (2) CSK

Actual Winners: (1) RR (2) CSK



Figure 10. Social mention insights for Match Date: 13/05/2018

11. **Match Date:** 14/05/2018

Matche:

- (1) KXIP vs RCB

Prediction based on Model: (1) RCB

Actual Winner: (1) RCB

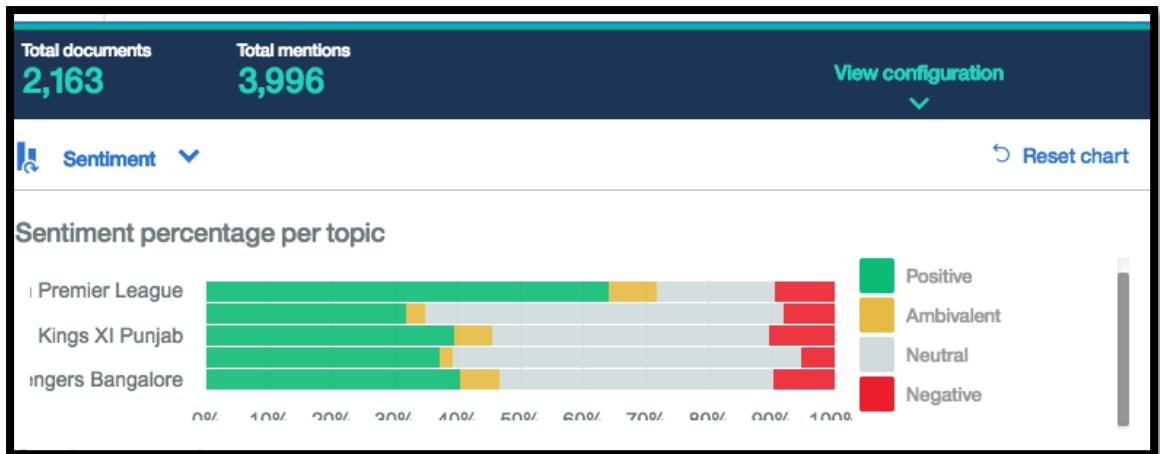


Figure 11. Social mention insights for Match Date: 14/05/2018

12. Match Date: 15/05/2018

Match:

(1) KKR vs RR

Prediction based on Model: (1) KKR

Actual Winner: (1) KKR

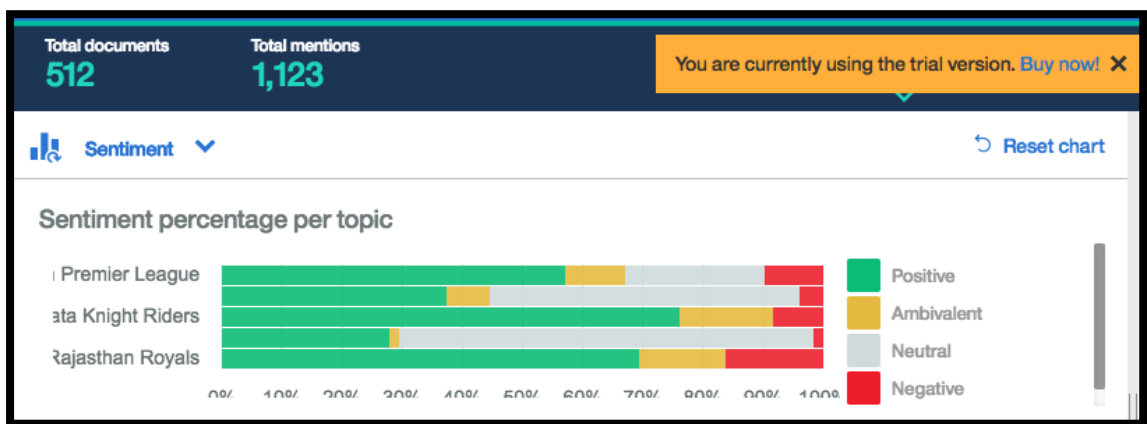


Figure 12. Social mention insights for Match Date: 15/05/2018

Based on the model, out of 20 predictions, 16 were right and 4 went wrong. Based on the same data, an SVM table was created as follows:

Table 1. Support Vector Machine (SVM) based prediction precision

Team	Matches	Won	Lost	Wins (Predicted)	Recall	Precision	F- Measure
Rajasthan Royals	5	3	2	3	0.6	1	0.75
Sunrisers Hyderabad	5	3	2	4	0.6	0.75	0.67
Delhi Daredevils	5	1	4	3	0.2	0.33	0.25
Mumbai Indians	5	2	3	1	0.4	2	0.67
Royal Challengers Bangalore	5	3	2	1	0.6	3	1.00
Kolkata Knight Riders	5	2	3	3	0.4	0.67	0.50
Chennai Super Kings	5	3	2	2	0.6	1.5	0.86
Kings XI Punjab	5	3	2	3	0.6	1	0.75

6. INFERENCE DRAWN

The SVM table showcases the precision in predictions. A high value of F-Measure shows the accuracy of the model for each of the teams listed about with RCB having the highest of 1. The high values also translate to the teams having higher chance to win the tournament (Mustafa et al. 2017). Based on the model we have tried to predict the result of match:

Match Date: 19 /05/2018

Matches:

(1) KKR vs SRH

(2) RR vs RCB

Prediction based on Model: (1) KKR (2) RCB

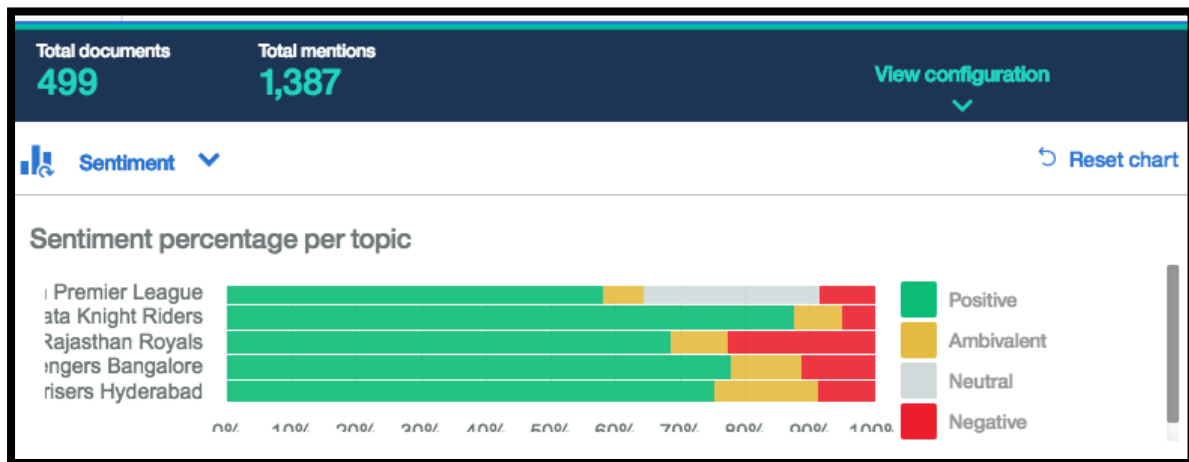


Figure 13. Social mention insights for Match Date: 19/05/2018

7. ACTIONABLE INSIGHT DRAWN

- Based on the prediction model, one could bet on the team in online betting sites similar to Dream 11 and attain an 80 % accuracy.
- A predictive model could be developed for forecasting the results to identify which brand or team to endorse for, if you are a brand. This would help in penetrating into the masses.

CONCLUSION AND RECOMMENDATIONS

The paper tries to map out the results solely based on the discussion happening in the social space, this could be utilised to monetize the results. For example, associations or partnerships with the winning team by predicting the winners before the match would help a firm to gain more revenue and brand value. The model could be expanded to other sports leagues, world cups and even Olympics. Similar model could be used to predict revenue of movies based on the sentiments before and after they release. One thing we should be careful about is the misuse of the same by teams to manipulate fake sentiments through fake social media accounts to gather more sponsors. Overall, the model proves to be ideal for a money-packed league like IPL.

REFERENCES

1. Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.

2. Ferrucci, D. A. (2012). Introduction to “this is watson”. *IBM Journal of Research and Development*, 56(3.4), 1-1.
3. Laroche, M., Habibi, M. R., & Richard, M. O. (2013). To be or not to be in social media: How brand loyalty is affected by social media?. *International Journal of Information Management*, 33(1), 76-82.
4. Mahyera, R. (2012) *Saving Cricket: a proposal for the legalization of gambling in India to regulate corrupt betting practices in cricket*. Emory International Literature Review.
5. Mustafa, R. et al. (2017) *Predicting The Cricket Match Outcome Using Crowd Opinions On Social Networks: A Comparative Study Of Machine Learning Methods*. *Malaysian Journal of Computer Science*. Vol. 30(1). pp 63-76.
6. Nyer, P. U. (2000). An investigation into whether complaining can cause increased consumer satisfaction. *Journal of consumer marketing*.
7. Shivhare, S. N., & Saritha, S. K. (2014). Emotion detection from text documents. *International Journal of Data Mining & Knowledge Management Process*, 4(6), 51.