

Date: September 11, 2017

Big Data & Analytics: The Next Big Thing In Elections

Businesses are using various methods in big data and analytics to analyze customer's data to improve competitiveness. Application of these methods not only generate vast amount of data but also result in unstructured information

Dr Monika Mittal



Currently, businesses are using various methods in big data and analytics to analyze customer's data to improve competitiveness. Application of these methods not only generate vast amount of data but also result in unstructured information. Further, to extract precise and structured data, Text Mining tools are applied. Text Mining or Data Mining is analysis of data contained in natural language text that can help an organization derive potentially valuable business insights from text-based content such as word documents, email and postings on social media streams like Facebook, Twitter, LinkedIn etc. (courtesy - <http://searchbusinessanalytics.techtarget.com/definition/text-mining>). It basically machine reads all the articles/ content full-text relevant to a topic, finds the right information and provides with what is required.

From Bio Medical to Academic research, security intelligence to Social media monitoring, organizations from diverse industry, across the globe, have adapted Text Mining Techniques and the results so far have proved to be efficient, cost reducing and improvisation in overall customer experience for businesses of every kind. Of all the sectors, Healthcare & Pharma, Academic, Automotive have benefitted the most from Text Mining application. They act as a bridge between structured and unstructured data and produce clusters of opinion in the business, helping in better decision making with real time feedback. Countries

such as US & Europe have started adapting this technique for polls during elections as well, as it examines overall sentiment of general public, analyzes trends and also a look into whether a particular party's opinion had any major impact or attitude towards a topic discussed over a time.

Outline of Polls taking place in India

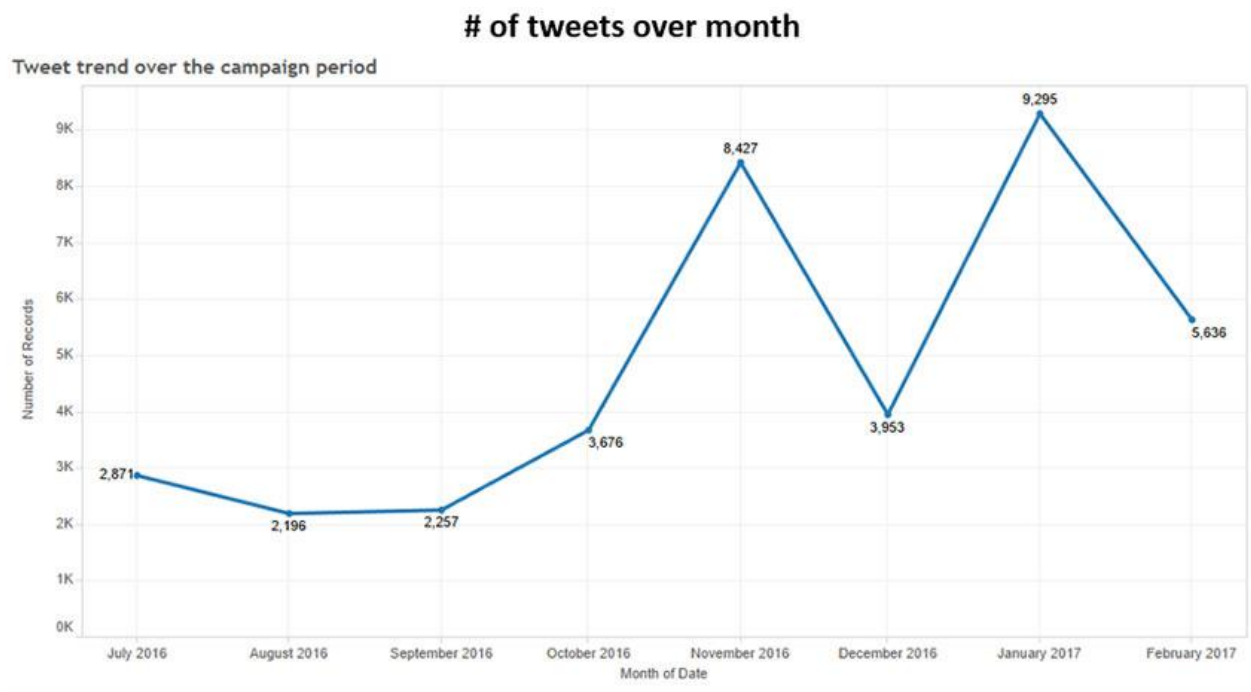
SCORE CALCULATION											
Election Issues	Ranking	Points	AAP			SAD			Congress		
			Positive	Negative	Score	Positive	Negative	Score	Positive	Negative	Score
Drug Issue	1	10	0.96	0.01	9.5	0.91	0.03	8.8	0.96	0.01	9.5
Unemployment	2	9.5	0.02	0.02	0	0.03	0.12	-0.855	0.04	0.03	0.095
Law and order	3	9	0.15	0.32	-1.53	0.09	0.66	-5.13	0.35	0.22	1.17
Communal Issues – Sacrilige of holy texts	4	8.5	0.03	0.07	-0.34	0.01	0.11	-0.85	0.03	0.14	-0.935
Water and Irrigation -	5	8	0.03	0.1	-0.56	0.02	0.14	-0.96	0.02	0.11	-0.72
Farmer Issues – Loan and Suicide	6	7.5	0.06	0.6	-4.05	0.04	0.03	0.075	0.15	0.34	-1.425
Corruption	7	7	0.39	0.14	1.75	0.21	0.5	-2.03	0.89	0.03	6.02
Women Safety	8	6.5	0.03	0.03	0	0.03	0.07	-0.26	0.03	0.02	0.065
Nepotism	9	6	0.3	0.27	0.18	0.87	0.07	4.8	0.91	0.02	5.34
Healthcare	10	5.5	0.08	0.02	0.33	0.02	0.08	-0.33	0.04	0.02	0.11
Education	11	5	0.41	0.11	1.5	0.1	0.11	-0.05	0.02	0.11	-0.45
NRI	12	4.5	0.06	0.02	0.18	0.06	0.11	-0.225	0.02	0.03	-0.045
Electricity	13	4	0.05	0.02	0.12	0.01	0.1	-0.36	0.08	0.02	0.24
Dalit Issues	14	3.5	0.01	0.01	0	0.02	0.1	-0.28	0.02	0.06	-0.14
Development – Infrastructure	15	3	0.03	0.03	0	0.02	0.08	-0.18	0.04	0.03	0.03
Border Tensions	16	2.5	0.03	0.05	-0.05	0.02	0.3	-0.7	0.05	0.02	0.075
Overall Score					7.03			1.465			18.93
			Popularity		0.26			0.05			0.69

Speaking of Elections & Polls, it is one of the most debatable & trending topic in the news in our country. Polls, especially, attract a great deal of attention for their ability to predict the outcome of elections. They are the talk of the town when elections are nearby as they play critical role, to an extent that they even influence voting preference of the undecided voters. Though exit polls announce the winner immediately, Opinion polls are the ones that boost up and give a fair picture of the scenario. Among other things, they help to point out issues that are important, how candidate qualities may affect voter's decision and how much support there is for a particular policy change by a party. But in recent times, Opinion polls have failed to predict results accurately, and have become less credible and misleading. And these issues arrive from various causes such as biased opinion of a poll agency, sampling error, using faulty demographic data and the model's inability to gauge public sentiment before the elections.

In the past, polling was the only way to collect data on who voted for who as the votes were casted anonymously. These approaches have become somewhat redundant in this digital age. Like other businesses and corporates, social media can be considered as a tool to develop a high cost efficient scientific method for poll prediction using Text Mining technique where we can pinpoint the issues that are important to different demographic groups, and identify how a candidate party is seen in regard to those

issues. This will also be of help to predict a winner based on the popularity index in conjunction with the campaign manifesto released by the political parties standing for elections.

How this can be achieved?



A paper work (at Great Lakes Institute of Management, Chennai) using Text Mining technique was done to analyze a political party's image on real time basis. We examined 1, 50,000 political tweets, which were published on Twitter's public message board between 2016-07 and 2017-02, prior to the Punjab state election in India, with increase in volume as the election drew nearer. We collected all the tweets that contained the names of any three parties represented in the Punjab State assembly. Considering Twitter as the social media polling tool and applying Text mining techniques to mine twitter data, We first analyzed the issues such as Drug usage, unemployment, water and irrigation etc., that were topmost in the electorate and identified how a candidate / party is seen in regard to those issues. A list of most critical issues using bag of words or running through various news websites, for the concerned region, where elections are to occur, were collected. The comments/tweets collected for each of the issues was segregated party wise and stored separately. The data was then analyzed and categorized as positive, negative or neutral sentiments.

Each campaign issue identified was given a weightage point based on the importance or criticality. The positive and negative sentiments are scored against each issue for each party considering the weightage point in calculation as well. Eventually an overall score for each party is obtained by calculating the score for each issue party-wise. This score is converted into percentage which provides us a popularity index for each party contesting in the elections. This might as well predict a winner based on the popularity index

calculated in conjunction with the campaign manifesto released by the political parties standing for elections.

Text Mining helped us analyze the sentiments posted over social media for major issues during elections. These comments were first extracted from the source and then collected for further work. Text mining proved to be a highly cost efficient scientific method for poll prediction where we were able to pinpoint the issues that are most important to the electorate.

Final Popularity Score

Score = (Positive Sentiment – Negative Sentiment) * Point

Final Popularity Score	
Party	Popularity %
AAP	26%
SAD	5%
Congress	69%

Though businesses are adopting this method at a slower rate, leaving a lot of room for the improvement. If parties begin using Text Mining techniques on a consistent basis, it will not only give a real-time picture of public image of the party but also provide insights to improve the reputation. The parties can design their campaign, deliver what the public needs, and do course corrections over the course of time.

Online link- <http://businessworld.in/article/Big-Data-Analytics-The-Next-Big-Thing-In-Elections-/11-09-2017-125853/>