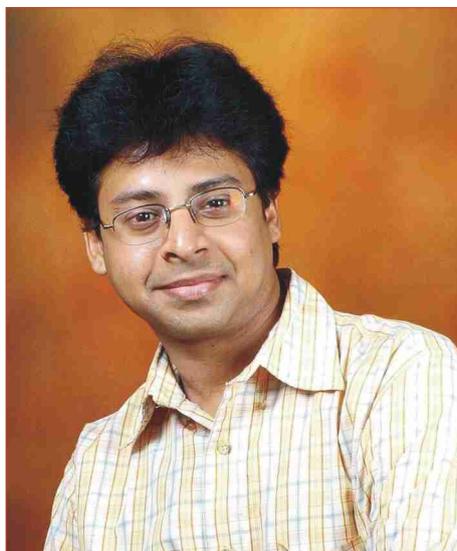


## BEST FEATURE IN IT JOURNALISM



### Srikanth RP

---

***Srikanth RP won the PoleStar Award for his article, 'How Indian TV show Satyamev Jayate used Big Data to inspire the world', which appeared in InformationWeek, dated August 2012.***

---

Srikanth RP is a vastly experienced communications specialist with more than 16 years of experience, with core expertise in writing about the business of Information Technology.

Srikanth is currently working with United Business Media (UBM) in the capacity of Executive Editor, InformationWeek (India). UBM is one of the world's largest media organizations, which owns brands such as InformationWeek, INTEROP, Network Computing and Dr Dobbs Journal. Srikanth is fully responsible for the InformationWeek (India) edition, leading a team of talented journalists. Prior to this, he was playing the role of Head-Conference for INTEROP (Mumbai) and Cloud Connect (Bengaluru), besides being responsible for InformationWeek (India).

In the past, Srikanth has headed the editorial operations of the Bombay bureau of Express Computer, one of India's leading IT magazines, where he was responsible for conceptualizing and creating a robust editorial plan that accurately captured the pulse of existing industry and technology trends.

He has also been associated with Patni Computer Systems, (now acquired by iGate), where he was a key member of the communications and thought leadership team responsible for creating a compelling value proposition, through focused written communications such as books, whitepapers etc. After Patni, he was associated with Capgemini India Consulting Private Limited, one of the world's premier IT consulting companies, where he headed the Internal Communications function.

# How Indian TV show Satyamev Jayate used Big Data to inspire the world

Highly popular television show, Satyamev Jayate, is a perfect example of how Big Data can be used to deliver maximum impact on society

Srikanth RP

August 27, 2012



On 6th May 2012, as curious Indians sat down in front of their television sets to catch the first episode of 'Satyamev Jayate' — a television series focused on social issues in India, there was nervousness in the air. Besides the producers who were obviously eager to gauge how viewers rated the show, there was another team in the background who was waiting diligently to execute one of its toughest projects. This team was from Persistent Systems who was given the task of analyzing messages generated from social media. The show's producers, in a first, wanted to analyze the messages that they would receive on social media to not only better plan future episodes, but also use the data to push for improved governance.

For Persistent Systems, who had done many Big Data projects earlier, this project was unique and different as the scale and type of data that would be received was difficult to comprehend. "We had no precedent in the industry which could help us build a system. The data could come from in different formats from a variety of different sources, such as Facebook, Twitter, websites, SMS polls and phone voice messages. This data needed to be analyzed immediately to understand the impact that the show had at an individual and society level. What made all this much more complex was the fact that we had no dry run," says Mukund Deshpande, Head BI and Analytics Competency, Persistent Systems. All Persistent Systems could do is gauge the type of response that could be expected. An intense marketing activity and an actor as popular as Aamir Khan hosting the show meant that the show was expected to witness huge viewership and participation. Additionally, as the show was based on social issues, it was expected that a large amount of information would be required to be collected.

While it had no operating guidelines, Persistent designed a system it believed was flexible enough to handle the load. "We did not know what to expect. We just said — let's get ready for the firehose," says Deshpande, referring to the huge avalanche of data that was expected from tweets and other platforms.

## Finding a relevant tweet in a flood of information

The huge popularity of the show surprised everybody. And then the flood of data began flowing in from all sources. The first episode on 'female foeticide' gathered 1.4 million responses from all sources. The data was in different formats — text, audio and video and in different languages that modern India was comfortable with — English and Hinglish (Mixture of Hindi and English). As the popularity of the show grew, so did the tweets and the messages from different social networks.

"We aggregated more than half a million tweets for season 1. We saw traffic of around 40,000 tweets on an average during the 90 minutes of the show and observed Twitter traffic used to be higher on Sunday and Monday. We followed a two-phased approach for the Live Analytics used during the show. We leveraged crowdsourcing for analysis and for deeper insights, we built algorithms to filter out the relevant tweets," explains Deshpande.

During the course of the engagement, there were close to 1,000 people divided into three groups involved at various stages of the development process. Persistent decided to build a software platform to filter/tag contents in a systematic manner. This tool was developed by assembling an array of automated tools to parse the data and a user interface for several analysts to process messages for deep analytics. The result is a cluster-based analysis along with trend, demographics and sentiment analysis for each message. The final step involved a manual check to find the latest and relevant top story. The results were aggregated and then further used for creating visualizations and dashboards. The analysis was done for all the 13 episodes over a thirteen week period. Persistent's challenge was building the right



**We saw traffic of around 40,000 tweets on an average during the 90 minutes of the show. We followed a two-phased approach for the Live Analytics used during the show – we leveraged crowd-sourcing for analysis and built algorithms to filter out the relevant tweets** ”

**Mukund Deshpande**  
Head BI & Analytics Competency,  
Persistent Systems



**1,249,440,319 Connections**



**14,972,514 Responses**



**8,115,739 Community Members**



**97,796,924 In Donations**



taxonomy for helping the system automatically sort out messages and unearth the right 'emotion' from the message. Messages included personal experiences, messages of hope, requests for help, solutions to problems, opinions about the topic and suggestions for new episodes. More than 50 tags were considered per episode for the analysis, and every message was scanned for these tags.

Sentiment analysis for a social topic is really complex. For example, given the topics addressed, such as female foeticide and medical malpractices, the tweets would be negative as it was more likely that the tweets used terms which were negative. But in reality, most people were positive and enthusiastic about the show. If a typical algorithm was used, most tweets were bound to be shown as negative. Persistent tweaked its algorithm on a constant basis, so that it could analyze and present the most relevant tweets.

### The Big Data problem

Given the huge volume of data that needed to be analyzed instantly and on a constant basis, Persistent Systems had a tough task on hand. For example, in season 1, the show created 14 million responses and made more than 1 billion impressions over the web. This was a typical Big Data problem and made the task of data analytics challenging.

Unlike other Big Data problems, in the case of 'Satyamev Jayate', the messages were varied and diverse. As the firm handled more and more data, Persistent kept on tweaking the system to deliver the best results. The team created dashboards for the producers and created the most meaningful visualizations. This information was used by the team to validate the appreciation, hypothesis and assumptions about the show. The data was also made available on the impact section of the show's website. When the Satyamev Jayate team spoke to the government leaders, the team from Persistent made sure that the team had the data to back its claims.

The detailed analysis of data coupled with the persistence of the producers, has ensured that the information collected does not

go to waste. For example, the Chief Justice of the Rajasthan High Court has agreed in principle to set up fast track courts in Jaipur to quickly resolve cases related to female foeticide. The Lok Sabha passed the child protection bill, which was pending for a long period of time.

### Using Big Data for governance

In the new era where everybody is connected with social or mobile connections, it is imperative that the government looks at these mediums to collect information. Today, most policy decisions are done using surveys that include only a small sample of respondents. This data is not representative of the whole country, and may be contained to a small region. By collecting and analyzing information, as Persistent has done, decisions and policy formulations can be done more accurately by using a larger set of data.

## Big Data Facts, Issues and impact @ Satyamev Jayate

- The show created 14 million responses and made more than 1 billion impressions over the web
- Persistent had to analyze data which came in different formats from a variety of different sources such as Facebook, Twitter, websites, SMS polls and phone voice messages
- Persistent deployed 1,000 people, which were divided into three groups and were involved at various stages of the development process
- The analysis of Big Data has created a huge impact. The Chief Justice of the Rajasthan High Court has agreed in principle to set-up fast track courts in Jaipur to quickly resolve cases related to female foeticide. The Lok Sabha passed the child protection bill, which was pending for a long period of time.