



# The Computerworld Honors Program

Honoring those who use Information Technology to benefit society

## Final Copy of Case Study

**Status:**

Laureate

**Year:**

2013

**Organization Name:**

Persistent Systems Ltd.

**Organization URL:**

[www.persistentsys.com](http://www.persistentsys.com)

**Project Name:**

Big Data/Social Media/Viewer Response Analytics for SMJ, a revolutionary new Indian TV show dedicated to resolving social issues

**Please select the category in which you are submitting your entry**

Emerging Technology

**Please provide an overview of the nominated project. Describe the problem it was intended to solve, the technology or approach used, how it was innovative and any technical or other challenges that had to be overcome for successful implementation and adoption. (In 300 words or less.)**

Persistent Systems developed a comprehensive analytics platform for Bollywood actor/film maker Aamir Khan's pioneering, interactive socio-cultural TV show in India, Satyamev Jayate/Truth Alone Prevails. The show, which tackled socially sensitive issues -- female foeticide, child sexual abuse, the caste system etc. -- had two purposes: 1. Create public debate about social issues. 2. Bring about change at the individual level. To achieve this, viewers needed to express opinions, share stories, comment about the issues. This created a Big Data problem with millions of unstructured responses submitted via Twitter, Facebook, YouTube, website comments, SMS, and voice messages, coupled with volume (15M responses, 1.1B impressions) and velocity (1M responses per episode). Persistent Systems designed a custom end-to-end analytics process in three weeks. The show was flooded with a staggering 1,098,786,391 (over a billion) impressions across social channels. All structured/unstructured data was analyzed in real-time to convey the show's impact on legislation (passing of child

protection act against sexual abuse), society (education campaigns on sexual abuse/nationwide debate on female foeticide) and the individual transparently shown on an impact dashboard: <http://www.satyamevjayate.in/impact/impact.php>. Adding to the unstructured data challenge, social media responses were in "Hinglish" (Hindi words in roman script embedded in English). This ruled out use of existing tools to handle messages. A new system was created to understand response sentiment. Deep analytics extracted valuable insights. The new system aggregated all unstructured data then automatically filtered data to weed out spam/unrelated messages. Valid messages were tagged and rated. Shorter messages were rated lower than longer messages/stories. Final selection was done manually using triangulation to determine top content, creating one of the most comprehensive viewer response analytics ever used for a TV show.

**When was this project implemented or last updated? (Please specify month and year.) Has it incorporated new technologies and/or other innovations since its initial deployment? (In 300 words or less.)**

The project was implemented from April 2012 to Aug 2012 during Season 1 of the TV show.

**Is implementation of the project complete? If no, please describe the project's phases and which phase the project is now in. (In 300 words or less.)**

The implementation of the project is complete. The project was implemented using the latest technology in distributed computing, Hadoop. Persistent Systems developed the crawlers and data aggregators used to fetch the data from various sources into the database. The data (user generated content) was then analyzed by analysts using a Content Filtering Ranking and Tagging System (CFRTS) developed by Persistent Systems. This data was then processed using Hadoop for analytics, and visualizations were created for easy consumption of the data by show viewers, the host, producers and researchers.

**Please provide at least one example of how the technology project has benefited a specific individual or organization. Feel free to include personal quotes from individuals who have directly benefited from the work. (In 300 words or less.)**

This revolutionary Big Data/Social Media Analytics project created the necessary digital feedback loop from viewers to viewers to the show host, the producers and research team. It helped viewers understand how people nationwide felt about issues, (which has since impacted legislation, education and individuals) while helping the show's producers understand viewership patterns, the impact of the show and comments/criticism. The production house (Amir Khan Productions) benefited from content analysis, suggestions for the next season's show and views from the other side, while the field research team got a 360-degree view of the social issues, quick insights into the public perception of the show and accurate dashboards during discussions with government policymakers. The project digitally empowered people to respond to social issues by sharing their own stories, solutions and stories of hope. The impact was felt at the government level when lawmakers listened to people's feedback on many pending resolutions. Finally, there was an impact at the individual level, where a viewer could think and respond about what



he/she could do about the social issue. "Persistent Systems collected more than a billion traces, which makes SMJ the most talked about show on earth. I asked the team one question: can you tell me in one sentence, what is the one thought that emerges for you from the millions of messages that have been flying around on the show? They thought about it and the one line the team has come back with is: I am not alone. That's a very big thought and also a very personal thought. Because people shared such intense personal stories on the show, the response has been very personal." Host and producer of the show, Amir Khan.

**Would this project be considered an innovation, a best practice or other notable advancement that could be adopted by or tailored for other organizations and uses? If yes, please describe that here. (In 300 words or less.)**

Today, many enterprises are struggling to understand how to use "Big Data" effectively with gigantic amount of data with increasing volume, variety (non-traditional unstructured sources like social media) and velocity (for example 400,000+ tweets/minute). It is becoming increasingly important to not only understand and analyze the user/customer response but to have them participate actively and achieve business goals. Big Data analytics can transform the relationship between users/customers and the business from a reactionary to an interactive mode. Comprehensive analytics can be used in vigilance, compliance or sentiment reporting to get a feel for the user/customer community. One-way communication is no longer enough; enterprises need active participation from users/customers. This type of analytics can be a critical factor in helping enterprises be more responsive and compete more effectively. Many large organizations today are already using social platforms for internal discussions and communication. Techniques pioneered here can be used very effectively across enterprises to have active employee participation in organizational issues, emergency responses or feedback on employee policy changes. Externally enterprises can harness the power of Big Data/Social Media Analytics to better engage with and solicit customer feedback -- as well as increase the enterprise's ability to quickly, effectively and efficiently address customer issues and concerns.

**If there are any other details that the judges should know about this project, please note them here. (In 300 words or less.)**

Appendix 1: Impact dashboard showing global impact 1B impressions, Appendix 2: Worldwide response map, Appendix 3: Example of the visual dashboard showing Twitter response over the globe. Media Articles about the project: a. Gigaom: <http://gigaom.com/cloud/how-indias-favorite-tv-show-uses-data-to-change-the-world/>, b. DataInformed: <http://data-informed.com/indian-tv-show-uses-social-media-analytic>.