

# BDA 2017 Tutorial Proposal - Application of Twitter in E-Governance

Swati Agarwal<sup>1</sup> and Ashish Sureka<sup>2</sup>

<sup>1</sup> Birla Institute of Technology & Science, Pilani, Goa, India,  
Email: [swatia@goa.bits-pilani.ac.in](mailto:swatia@goa.bits-pilani.ac.in)

<sup>2</sup> Ashoka University, India,  
Email: [ashish.sureka@ashoka.edu.in](mailto:ashish.sureka@ashoka.edu.in)

**Abstract.** Open Source Social Media Intelligence (OSSMInt) is a field that focuses on extracting useful information and actionable insights from publicly available and overt sources of data on social media platforms. There are several applications that can be built by applying OSSMInt techniques on this human-sensor data. In this tutorial, we will present some of the use-cases of OSSMInt that are useful for the public sector agencies for e-governance. E-governance on social media include the identification of complaints and grievances reported online by the public citizens for the government authorities and facilitate public agencies to response those complaints, provide better services and improve their connections with public citizens. We will introduce the basic Natural Language Processing and Machine Learning based framework, tools and techniques within the context of OSSMInt and E-governance. The focus of our tutorial is on mining user-generated data on Twitter. In particular, we will focus on two important applications: 1) complaints which are reported to spread awareness among other citizens and to bring government's attention to the issues reported in the complaint, and 2) complaints which seek for immediate action and response from the concerned authorities. In addition to the basic introduction and motivation, we will discuss the unique challenges to these applications, open research problems, important literature, proposed approach, experimental results, and future directions.

**Keywords:** Bad Roads Complaints, Complaints and Grievances, Government Applications, Information Visualization, Lexical KnowledgeBase, Mining User Generated Content, Natural Language Processing, Social Media Analytics, Text Classification, Twitter

## 1 Basic Information

**Duration** Half-Day (3-4 Hours)

**Pre-requisite** Basic understanding of Social Media and exposure to data mining tools and techniques

**Target Audiences** MS/MTech and PhD students, Faculty Members and Researchers in private and public sector, working or interested in the area of Social Media Analytics, E-governance, and Organization-Consumer Management

**Learning Outcome** The tutorial will cover fundamental techniques, applications, research problems and future directions. Following are the 3 specific learning outcome:

1. Familiarity with E-governance applications (complaints and grievances mining) in the domain of Social Computing and Social Informatics
2. Applications of Natural Language Processing, Machine Learning and Data Mining based framework, tools and techniques for citizensourcing
3. Overview of important research problems in the area of e-governance on social media, solution approaches, results and conclusions from recent research papers.

**Previous BDA Tutorial Presentation Experience** We presented an invited tutorial (Duration: half-day) at BDA 2015 held in Hyderabad, India. The topic of our tutorial was "Open Source Social Media Analytics for Intelligence and Security Informatics Applications" [6]. The presentation went well and was attended by more than 30 people including graduate students, and researchers from academia as well as industry. Link to our tutorial paper: [https://link.springer.com/chapter/10.1007/978-3-319-27057-9\\_2](https://link.springer.com/chapter/10.1007/978-3-319-27057-9_2)

## 2 Tutorial Outline

The area of developing applications for e-governance by using social computing is a multi-disciplinary area. The scope of this tutorial is an intersection of (1) Online Social Media Platforms (2) Government and Public Agencies (3) Machine Learning and Natural Language Processing techniques. In particular, we will discuss mining user generated content on social media websites to identify public complaints and grievances. The focus of the tutorial will be to present basic introduction of usage of social media in E-governance and two case studies on identification of citizens' complaints and grievances posted on social media. The tutorial will be based on some of the recent research work and publications by the authors [1][2][7][11][3][4][5]. Following is the list of **topics** and **case-studies** to be covered in the tutorial:

1. Web 2.0 and Social Media Platforms
2. Open-Source Intelligence (OSINT)
3. Open-Source Social Media Intelligence Applications
4. Machine Learning Framework
5. Mining Social Media Textual Data
6. Case-Study on Identification of public complaints to Ministry of Railway, Traffic, Income Tax and State Police.
7. Case-Study on automatic classification of public complaints reported on bad road condition.
8. Open Research Problems and Future Directions

## 2.1 Identification of Public Citizens' Complaints on Twitter

We formulate the problem of automatic identification of citizens' complaints (reported to official Twitter handle of a public agency) as a one-class classification problem. We propose a text classification based approach consisting of various components performing several tasks: tweets extraction from public agencies' account, enrichment and enhancement of raw microposts (tweets), learning the features of non-complaint and complaint report tweets, developing a baseline classification approach, use of ensemble techniques to improve the baseline method, empirical analysis and performance evaluation. Furthermore, we will discuss the importance of visualization of empirical results to extract useful information and insights from the complaint reports.

## 2.2 Mining Twitter to Extract Information on Killer Roads

In this part of the tutorial, we will present our solution approach for identifying complaints and grievances reported on the bad road conditions in India. We will investigate the efficiency of natural language processing based techniques to identify the problem or issue reported in the complaints. Furthermore, we will demonstrate the importance of location and topic of the issue to identify the bad road (referred as killer road) conditions. We will present our solution approach consisting of various components such as raw text enrichment, linguistic features extraction, location verification, AISP (appreciation, information sharing and promotional tweets) classification, complaint report classification. We conduct a series of experiments on a real-world and large dataset and demonstrate the effectiveness of our approach. Furthermore, we will discuss the method to enrich the missing information in incomplete and nearly-useful complaint reports.

## 3 Presenters' Brief Bio

**Swati Agarwal**<sup>3</sup> is a Visiting Assistant Professor in Computer Science Department at BITS Pilani- Goa, India. Her research interests are in the area of Social Computing, Natural Language Processing, Security Informatics, and Text mining and Analytics. She has a PhD in Computer Science (Social Media Analytics and Security Informatics) from IIIT-Delhi. She has guided several undergraduate and postgraduate students in their independent research projects.

**Ashish Sureka**<sup>4</sup> is an associate professor at Ashoka University, India. His research interests are in the area of Social Media Analytics and Mining Software Repositories. He has a PhD in Computer Science from NCSU and has worked at IBM Research (USA), IIIT Delhi, Infosys, TRDDC, Siemens and ABB Corporate Research. He has graduated several PhD and MTech students.

---

<sup>3</sup> <http://www.swati-agarwal.in/>

<sup>4</sup> <http://www.ashish-sureka.in>

## References

1. Agarwal, S.: Extended abstract: Open source social media intelligence for enabling government applications. ACM SIGWEB Newsletter (Summer) (2017)
2. Agarwal, S., Mittal, N., Sureka, A.: Potholes and bad road conditions- mining twitter to extract information on killer roads. In: The ACM India Joint International Conference on Data Science and Management of Data (CoDS-COMAD), Goa, India. ACM (2018), under-review
3. Agarwal, S., Sureka, A.: Investigating the potential of aggregated tweets as surrogate data for forecasting civil protests. In: Proceedings of the 3rd IKDD Conference on Data Science, 2016. p. 8. ACM (2016)
4. Agarwal, S., Sureka, A.: A collision of beliefs: investigating linguistic features for religious conflicts identification on tumblr. In: Distributed Computing and Internet Technology, pp. 43–57. Springer (2017)
5. Agarwal, S., Sureka, A.: Investigating the dynamics of religious conflicts by mining public opinions on social media. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 421–433. Springer (2017)
6. Agarwal, S., Sureka, A., Goyal, V.: Open source social media analytics for intelligence and security informatics applications. In: International Conference on Big Data Analytics, pp. 21–37. Springer (2015)
7. Agarwal, S., Sureka, A., Goyal, V.: Open source social media as sensors for enabling government identification, prediction and response applications. Ph.D. thesis, IIIT-Delhi (2017)
8. Frias-Martinez, V., Sae-Tang, A., Frias-Martinez, E.: To call, or to tweet? understanding 3-1-1 citizen complaint behaviors. In: ASE Big-Data/SocialCom/CyberSecurity Conference (2014)
9. Khan, G.F., Swar, B., Lee, S.K.: Social media risks and benefits a public sector perspective. Social Science Computer Review 32(5), 606–627 (2014)
10. Meijer, A.J., Torenlvied, R.: Social media and the new organization of government communications an empirical analysis of twitter usage by the dutch police. The American Review of Public Administration p. 0275074014551381 (2014)
11. Mittal, N., Agarwal, S., Sureka, A.: Got a complaint?-keep calm and tweet it! In: Advanced Data Mining and Applications: 12th International Conference, ADMA 2016, Gold Coast, QLD, Australia, December 12-15, 2016, Proceedings 12. pp. 619–635. Springer (2016)
12. Panagiotopoulos, P., Barnett, J., Bigdeli, A.Z., Sams, S.: Social media in emergency management: Twitter as a tool for communicating risks to the public. Technological Forecasting and Social Change 111, 86–96 (2016)
13. Schulz, A., Ristoski, P., Paulheim, H.: I see a car crash: Real-time detection of small scale incidents in microblogs. In: Extended Semantic Web Conference. pp. 22–33. Springer (2013)