

Big Data Analytics: Towards a Model to Understand Development Equity for Villages in Indonesia

Annas Vijaya^{1,*}, *Linda Salma Angreani*², and *Mokhamad Amin Hariyadi*²

¹Information System Department, Faculty of Science and Technology, Universitas Ma Chung Malang, Villa Puncak Tidar N-01, Karangwidoro, Dau, Malang, East Java, 65151, Indonesia

²Informatics Department, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Jl. Gajayana 50, Malang, East Java, 65144, Indonesia

Abstract. The aim of this paper is to design a prototype model that can be used to better understand development equity for villages in terms of public monitoring and evaluation. In designing the model, the research has reviewed several techniques of big data analytics as well as alignment of business strategic objectives and technology. The prototype model also tested using several types of data. Although some obstacles have found, as it also found in the reviewed literature, a prototype model which can guide researchers and practitioners to understand ways to capture public monitoring is presented in this paper. Furthermore, Information systems researchers could use this prototype model for further research to get a deeper understanding of big data analytics roles for development, particularly in developing countries.

Key words: Big data analytics, big data for development, developing countries, villages.

1 Introduction

The village government in Indonesia is a nearest unit to society and involved directly in government programs for several basic development areas. Because of its importance, the Indonesian government has established a specific regulation through the Act. No. 6, 2014 [1] about village governance. Nevertheless, there are issues in order to control village developments which was managed by village government. Reports of village misappropriation, including corruptions in the use of village funds are still occurred [2]. By using IT as an enabler, public control of village development, particularly from the village community over their own village, is expected to create a positive contribution to village development in Indonesia. A technology that have potentials to meet this expectation is big data analytics.

Despite big data analytics perceived have strategic potential benefits for developing countries [3], there are some challenges in implementation phase, such as data access and privacy, human resource capacity, IT infrastructure, IT management, and financial capabilities [4–7]. Furthermore, some developing countries are facing issues in digital divide and analytical processes such as methodology, interpretation accuracy, analytical methods and anomaly detection [6, 7]. This has motivated this study to further explore the creation of the prototype model to address those challenges.

In the attempt to address the situation, this paper is constructed in several sections as follows. Section 1 described background and issues related to the topic. Section 2 represented several related studies which highlight possible technologies related to big data for development and techniques of big data analytics. Next section, Section 3, described a way to address those problems by designing a prototype model that captured specific public information, and as a tool to support public control over village development. Before developing a model, in order to create alignment between strategic objectives and technology, we mapped several strategic objectives as stated in the Act (Act. No. 6, 2014) to technology, in this case is big data analytics [8]. This process included mapping the objectives to critical success factors, critical information, and solutions of big data analytics. The model presented in Section 3 is tested using several types of data which are call data record (CDR) and Twitter data. From several previous studies, the use of this service model could gain more benefits by presenting real-time information [3, 9]. Following this section, Section 4, results from the study to address the challenges is presented and discussed. Finally, a summary will ends this paper as a final section.

*Corresponding author: annas.vijaya@machung.ac.id

2 Literature review

The use of big data analytics has several potential areas for development in the developing world. Refer to results of Bellagio big data workshop in 2014, which consist of several activists, researchers and data expert; those potential areas including advocating and facilitating; describing and predicting; facilitating information exchange and promoting accountability and transparency [3]. Later, the term used for these purposes is big data for development which coined by UN Global Pulse [7].

Recently there are several research and applications related to big data for development using various data sources such as social media [9], cell phone usage [10–12], digital transactions [13], online news media [9, 14], and administrative records [15]. Some of these applications are using public datasets which were become critical elements of big data for development [7]. The combination of data capture process and information management on those studies are used in the prototype model.

The data analysis used in this paper refers to big data analytics as part of business intelligence and analytics (BIA) which become an important part for researchers and industries in more than two decades [16–18]. Big data analytics is described as in-app analysis techniques with a large and complex set of data sets that require unique data storage, management, analysis, and visualization technologies [3, 17, 19], which also used in the prototype model in this paper.

3 Prototype design

Figure 1 shows the design of prototype model. As previously mentioned, there are two types of data used, which are short message system (SMS) data from CDR and Twitter data. There are several reasons to use these types of data. Firstly, the biggest active Twitter users from developing countries is came from Indonesia [20]. This fact shows that netizens in Indonesia are often express their opinions and critiques on actual issues including those related to village development using social media. Secondly, although social networking can be used as a source of data, there are issues related to digital divide in Indonesia, and the evidence shows that Twitter data are only valid in major cities. Therefore, we used SMS data from CDR as second sources of data as it used more widely in rural areas in Indonesia. Finally, these two types of data are considered quite representative for the prototype model, however, it is possible to use other types of data as long as it already filtered as a clean data.

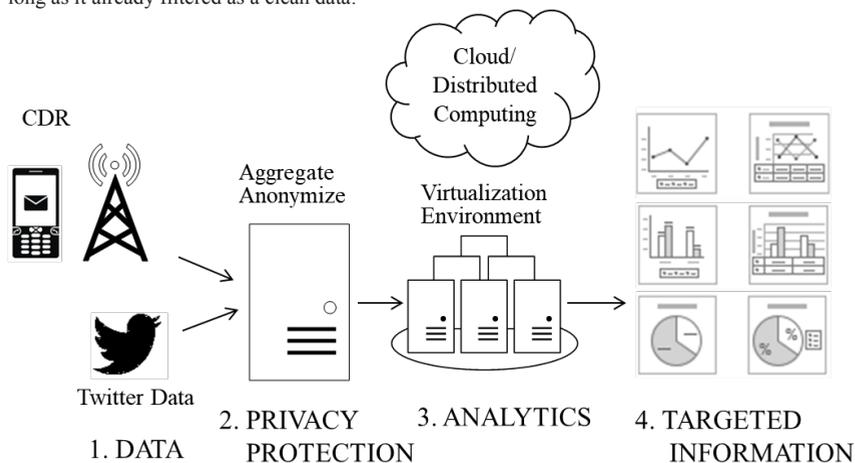


Fig. 1. Steps of prototype design.

After retrieving those type of data, next process is anonymize process as part of privacy protection process. Anonymizing CDR data is done by deleting cellular ID, while user ID was erased in the Twitter data. Following this step is the analysis process as shown in Figure 2, which includes information analysis and topic models. There are at least two experimental environment models with different methods for testing which are virtualization and physical cluster. However, in real situation there will be a third environment which is cloud computing using the same technical methods as in the physical cluster model. Figure 3 shows these three models.

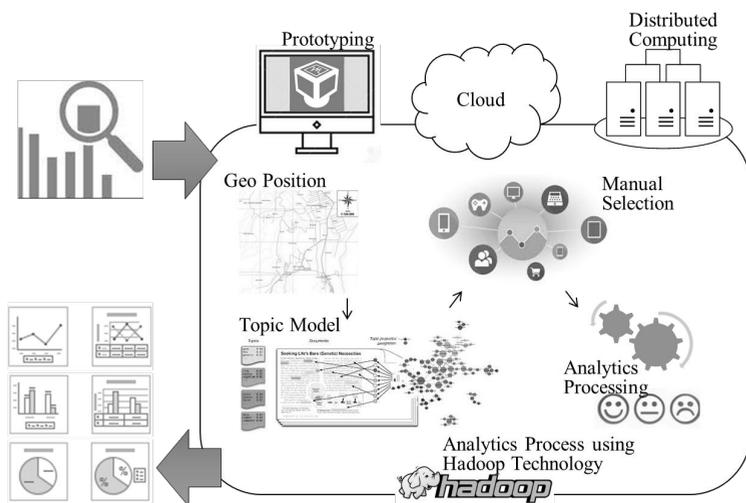


Fig. 2. Steps of big data analytics.

4 Results

As mentioned before, we have done the mapping process of business objectives to technology we used. Table 1 shows the results of this mapping process. After getting the result, we made a classification of needs of big data analytics. The classification result shows that prototype model is directed to provide the following information:

- (i) Information related to the management of village government, which includes:
 - i). Monitoring and reporting of village existence (1A);
 - ii). Transparency of village government monitoring and reporting, including village development activities, development of village potentials, economic activities, and use of village funds (1B);
 - iii). Monitoring and reporting of village public service (1C).
- (ii) Information of village community activities, including:
 - i). Social and cultural, such as conservation and promotion of village customs, traditions, and culture (2A);
 - ii). Activities of village resilience and village politics (2B)

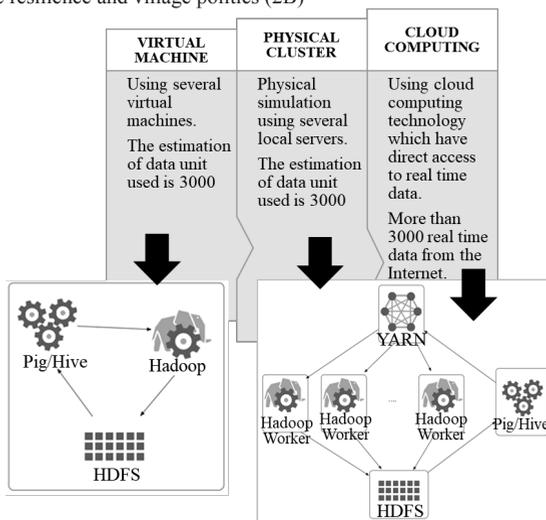


Fig. 3. Experimental environment model for the prototype.

Table 1. Mapping the objectives to achieve the alignment of technology.

| Objectives of Village Management based on Government Regulation (UU No.6, 2014) | CSF on IS/IT Domain | Solution of Big Data Analytics |
|---|---|--|
| i). Provide recognition and respect for existing villages with their diversity before and after the establishment of the Archipelagic State of The Republic of Indonesia; | The existence of the village is recognized | Availability of information on monitoring and reporting of village existence |
| ii). To provide clarity on the status and legal certainty of the village in the constitutional system of the Republic of Indonesia in order to realize justice for all Indonesian people; | Clarity of legal status of the village | Availability of information on monitoring and reporting of village existence |
| iii). Preserve and promote the customs, traditions and culture of the village society; | The existence of efforts to preserve and promote the customs, traditions and culture of village communities | Monitoring and reporting activities for the preservation and promotion of customs, traditions and culture of village society |
| iv). Encourage initiatives, and participation of village societies for the development of village potentials and assets for mutual prosperity; | The existence of village community activities for the development of village potential and assets | System of monitoring and reporting of potential village development activities |
| v). Establishing a professional, efficient and effective, open and accountable village government; | The development of a professional, efficient, effective, open and accountable village government | A transparent village government monitoring and reporting system |
| vi). Improving public services for village societies to accelerate the realization of general welfare; | Achievement of accountable public services in the village | Monitoring and reporting system of public services in the village |
| vii). Improve socio-cultural resilience of village communities in order to realize village communities that are able to maintain social unity as part of national security; | The existence of activities to improve the social resilience of village culture | Monitoring and reporting system improving village resilience |
| viii). Promote the economy of rural communities and overcome the national development gap; and | The existence of activities that can promote the economy of rural communities | System of monitoring and reporting of economic activities and or use of village funds for village development |
| ix). Strengthen the village society as the subject of development. | The existence of activities that make village society as the subject of development | System of monitoring and reporting of village development activities |

Based on the analysis and design results, data capture process from data source began. Initial data capture from Twitter are done for 5 d with the initial query word 'desa' (which means village). Total initial data obtained from Twitter with geo-location Indonesia is 28 000 units. On the contrary, we have difficulties to access CDR data from telco operators. The cellular service providers who owned the data were bound by the government policy that the data access was prohibited. So we only get five different types of CDR data including calling and SMS data. We utilized those data to extract the structure and create a model of data filters on the prototype model.

Testing process shows us that process of anonymize data from Twitter takes up to 180 min. We also found that during data capture process, particularly in rural area, there is a possibility of lack of connection to the Internet. This indicates that infrastructure availability and reliability become crucial in the actual implementation. After finishing the anonymized data process, result data are converted into text form for classification of information using topic models

based on latent dirichlet allocation (LDA) [21], and the functions we used is based on online learning for latent dirichlet allocation [22]. Some modifications are made in order to make this functions suitable for the research purpose.

Result of LDA shows that several topics are match to the information classification in analytics process. Figure 4 shows percentage of topics that fit to the result of Table 1. Nevertheless, 32 % of data were not relevant to the strategic objectives. Therefore, carefulness in work while filtering the data is consider necessary. The final results of the information from prototype model can then be directed to specific information needs, such as public sentiment on village development as those undertaken by UN Global Pulse to understand food price crises [9], as well as other purposes summarized in the Bellagio big data workshop [3].

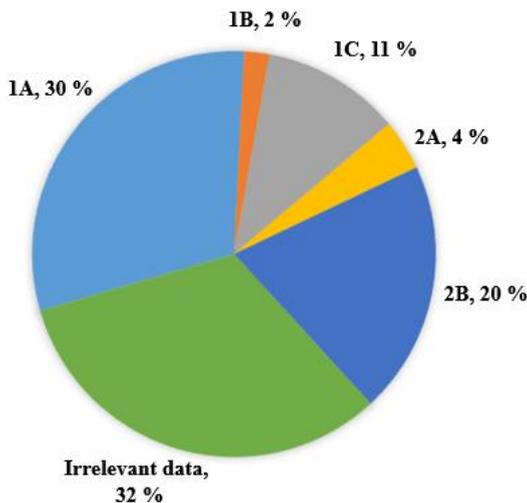


Fig. 4. Percentage of topics that fit to the result of the mapping objectives.

5 Conclusion and future work

Research conducted as described in this paper shows that there are obstacles in using big data analytics for villages in Indonesia such as data access, data privacy, IT infrastructure, digital divide, and anomaly detection. In spite of those obstacles, we found a relation between village management objectives as regulated by the Government of Indonesia and the topic of public discussion in the Internet as we captured the data. It shows us that big data analytics solution could be set as a tool to understand development equity for villages which provide public control and evaluation functions. However, it is important to note that another future work of big data analytics for villages in Indonesia will need higher-level exploration and longer research periods to obtain stronger evidence in terms of validity and effectiveness. The prototype model can also compare and combine with current model of monitoring and reporting of village government. For instance, the Government of Indonesia has provided a model of manual reporting system. By comparing and combining both, analytics result and reporting information, we can get a new perspective of the report in terms of public monitoring and evaluations. It would be easier if the adoption of domain: *desa.id* (a domain type for village in Indonesia) was utilized for both model. This may lead to another model to get deeper understanding of equitable development for villages.

References

1. Ministry of Justice and Human Rights of Republic of Indonesia. *Undang-undang nomor 6 tahun 2014 tentang desa*. [Law Number 6 Year 2014 About Villages]. Kementerian Hukum dan Hak Asasi Manusia Republik Indonesia, Jakarta, 2014. [in Bahasa Indonesia]. <http://www.jdih.kemenkeu.go.id/fullText/2014/6TAHUN2014UU.HTM>
2. M.P. Istman, I. Maulidar. *Gawat, penyelewengan dana desa makin berjibun*. [Disastrous, the misuse of village funds has become more and more overwhelming]. [Online] from <https://fokus.tempo.co/read/1000406/gawat-penyelewengan-dana-desa-makin-berjibun> (2017). [Accessed on 23 August 2017]. [in Bahasa Indonesia].
3. Bellagio Big Data Workshop Participants. *Big data and positive social change in the developing world: A white paper for practitioners and researchers*. Oxford: Oxford Internet Institute (2014).

- <https://assets.rockefellerfoundation.org/app/uploads/20140507222100/Big-Data-and-Positive-Social-Change-in-the-Developing-World.pdf>
4. J. Piotrowski. *Big obstacles ahead for big data for development*. [Online] from <http://www.scidev.net/global/data/feature/obstacles-big-data-development.html> (2014). [Accessed on 23 December 2015].
 5. UN Global Pulse. *Agile Global Development: Harnessing The Power of Real-time Information*. New York: UN Global Pulse (2011). <https://www.hashdoc.com/documents/2018/agile-global-development>
 6. ECOSOC. *Big data and modernization of statistical systems, report of the secretary-general*. United Nations, Economic and Social Council, Statistical Commission (2014). <https://unstats.un.org/unsd/statcom/doc14/2014-11-BigData-E.pdf>
 7. United Nations Global Pulse. *Big data for development: A primer*. [Online]. From http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf (2013). [Accessed on 24 January 2016]
 8. J. Ward, J. Peppard. *Strategic planning for information system, 3rd ed*. New York: John Wiley & Sons (2002). <https://www.wiley.com/en-us/Strategic+Planning+for+Information+Systems%2C+3rd+Edition-p-9781118585252>
 9. UN Global Pulse. *Mining Indonesian tweets to understand food price crises*. Jakarta: Pulse Lab Jakarta (2014). <http://www.unglobalpulse.org/sites/default/files/Global-Pulse-Mining-Indonesian-Tweets-Food-Price-Crises%20copy.pdf>
 10. D. Talbot. *10 breakthrough technologies 2013: Mobile data from cheap phones*. [Online]. From <http://www.technologyreview.com/featuredstory/513721/big-data-from-cheap-phones/> (2013). [Accessed on 5 December 2016].
 11. X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki. *Prediction of human emergency behavior and their mobility following large-scale disaster*. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 24–27 August (New York, USA, 2014). pp. 5–14. <https://dl.acm.org/citation.cfm?id=2623628>
 12. L. Bengtsson, X. Lu, A. Thorson, R. Garfield, J. von Schreeb. *PLoS Medicine* **8**(8):1–9 (2011). <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001083>
 13. World Economic Forum. *Big data, big impact: New possibilities for international development*. Geneva: The World Economic Forum (2012). http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf
 14. USAID Development Credit Authority, UN Global Pulse. *Feasibility study on the potential to gain insights on Kenyans' access to finance through analysis of big data*. New York: United Nations Global Pulse (2013). <https://usaidealmlab.org/sites/default/files/resource/files/UNGlobalPulse-USAID-DCA-FeasStudyFinalReport-27Sept2013.pdf>
 15. United Nations Global Pulse. *Annual Report 2013*. New York: UN Global Pulse (2014). http://www.unglobalpulse.org/sites/default/files/GP%20Annual%20Report_2013.pdf
 16. E.-P. Lim, H. Chen, G. Chen. *ACM Transactions on Management Information Systems*, **3**(4):1–10 (2013). http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2965&context=sis_research
 17. H. Chen, R.H.L. Chiang, V.C. Storey. *MIS Quarterly*, **36**(4):1165–1188 (2012). <https://ai.arizona.edu/sites/ai/files/MIS611D/chen-bi-december-2012.pdf>
 18. T.H. Davenport. *Competing on analytics*. *Harvard Business Review*, **84**(1):98–107 (2006). <https://hbr.org/2006/01/competing-on-analytics>
 19. P.B. Goes. *Big Data and IS Research*. *MIS Quarterly, Editor's Comments*, **38**(3):iii–viii (2014). <https://misq.org/misq/downloads/download/editorial/603/>
 20. V. Lipman. *The world's most active twitter city? you won't guess it*. [Online]. from <http://www.forbes.com/sites/victorlipman/2012/12/30/the-worlds-most-active-twitter-city-you-wont-guess-it/> (2012). [Accessed on 1 April 2016].
 21. M.D. Blei. *Probabilistic topic models*. *Communication of The ACM*, **55**(4):77–84 (2012). <https://dl.acm.org/citation.cfm?id=2133826>
 22. M. D. Hoffman, F. R. Bach and D. M. Blei, "Online Learning for Latent Dirichlet Allocation," in *Neural Information Processing Systems 23*, New York, (2010). <https://papers.nips.cc/paper/3902-online-learning-for-latent-dirichlet-allocation.pdf>