Big Data Analytic:Smart City

Srishti Yadav

Department of Computer Application, Babu Banarasi Das University, Lucknow, India
E-mail: srishtiyadav1245@gmail.com

Abstract:

The concept of smart cities has gained a lot of attention recently. The emergence of technologies like the Internet of Things (IoT), cloud computing, and big data analytics has made smart cities a reality. In this essay, we look at how big data analytics is used in initiatives that attempt to create smart cities. We discuss the many applications of big data analytics in smart cities, such as waste management, energy efficiency, and traffic management. We also examine the problems that big data analytics in smart cities raises in terms of data quality, data integration, and data privacy and security. Finally, we offer a number of case studies of effective big data analytics applications in smart cities as well as some suggestions for further research.

Keywords: big data analytics; smart cities; data-driven decision making;

Introduction

Smart cities will exist in the future. They are designed to provide its residents a high quality of life by applying technology to improve several areas of urban living. Smart city initiatives are based on the use of cutting-edge technology like the Internet of Things (IoT), cloud computing, and big data analytics. Because it enables the collection, analysis, and analysis of enormous volumes of data created by multiple sources, big data analytics is an essential component of smart city initiatives.

Smart city applications of big data analytics:

Big data analytics may be applied in many different ways in smart cities. The control of traffic is one of the most crucial applications. By evaluating data from many sources, like traffic cameras, GPS, and social media, big data analytics may help to lessen traffic congestion. In many places, traffic congestion is a significant problem. By evaluating this data, city officials can pinpoint areas that have the most traffic congestion and develop solutions to ease it.

Another area where big data analytics is applied in smart cities is energy efficiency. As part of smart city initiatives, energy utilisation in structures, lights, and other infrastructure is improved with the intention of reducing consumption. Big data analytics may help in accomplishing this goal by recognising trends in energy use and developing strategies to minimise it by analysing data from smart metres, sensors, and other sources.

Smart city waste management may also make use of big data analytics. By assessing data from many sources, such as sensors, social media, and weather predictions, city officials may develop strategies to maximise garbage collection routes, reduce waste output, and enhance recycling rates.

Big Data Analytics in Smart Cities: The Challenges

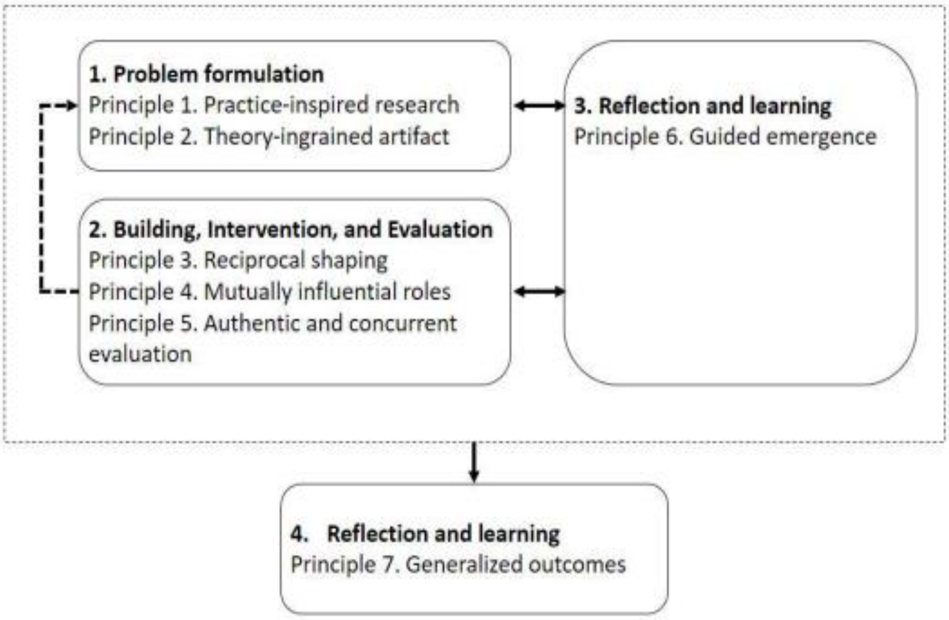
Big data analytics may help alter smart cities, but there are also numerous challenges involved. One of the major challenges is maintaining data privacy and security. Smart cities generate a lot of confidential and delicate data. The protection of this data from misuse and unauthorised access is the responsibility of the city authorities.

Data quality is another problem with big data analytics in smart cities. Data collection that is accurate and reliable is crucial for smart city initiatives. However, a variety of factors, such as sensor malfunction, data loss, and data corruption, might affect the data's quality. Local government must uphold data quality in order to ensure the success of smart city initiatives.

Data integration is a problem with big data analytics in smart cities. Smart cities generate data from various sources, such as sensors, social media, and official databases. City officials must make sure that this data is integrated and examined effectively in order to offer insights that can be put to use.

2. Research Method

We employed two techniques for this inquiry. We began by performing an extensive review of the literature on big data analytics frameworks for smart cities. The proposed publishing list is then scrutinized in relation to the listed SC domains. The objective of this stage is to review the most recent studies that offer SCs Big Data Analytics frameworks. The examination of the joint research directions for these two disciplines is aided by this survey. At the same time, it enables comparison of domain-agnostic Big Data Analytics frameworks in the context of Smart Cities in order to identify their commonalities. The end of this phase is marked by the selection of a framework that facilitates decision-making with the ability to share information. Utilizing a prototype of a certain Big Data Analytics framework is part of the second phase. To construct and assess the framework prototype while preserving a balance between organizational and technological dominance, we use the action design research (ADR) technique. Each of the four stages of ADR is founded on a group of seven guiding concepts.



Action design research stages and principles

Stage 1: Problem Formulation

For a topic that is believed to exist in practice or has been anticipated by academics, the initial research questions are established in this section. Utilizing contemporary theories and technologies, the research opportunity's concept was established. This step makes use of the two distinct notions of practice-inspired research and theory-ingrained artefact.

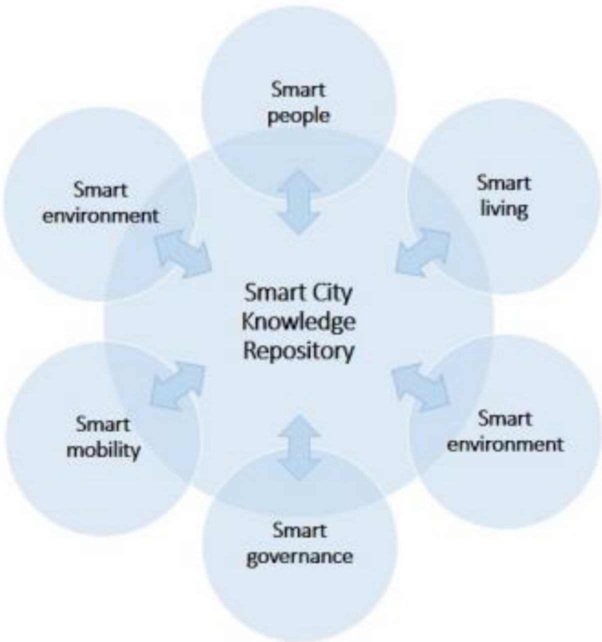
The introduction and background portions of this article both reflect this level. The theoretical and practical significance of the research question, "How can Big Data Analytics be used as a data-driven decision-making enabler in Smart Cities?" is established in these parts. The area of study that will be filled by responding to this question is highlighted in the literature review in Section 4. It is essential to frame the current problem as an illustration of a larger class of problems.

Stage 2: Reflection and Learning

This stage emphasizes that deliberate reflection on the theories chosen, the issue framing, and the evolving ensemble artifact is necessary for ensuring that contributions to knowledge are acknowledged. It is also essential to change the study strategy in light of preliminary evaluation results in order to accommodate for the evolving understanding of the ensemble artifact. Only the sixth principle, directed emergence, is used to describe this stage. In Section 8, which examines the design principles and generalizability of the artifact in the context of stakeholder feedback analysis, this process is illustrated.

3. Importance of the Research

From both an academic and an empirical perspective, this work is important. It first offers a new avenue of research into SC applications of BDA frameworks. Second, a novel idea sets SCDAP apart from previous BDA frameworks from a practical perspective: building a repository to hold the data model and extracted analytics that result from the analysis process. SCDAP is a domain-neutral framework that may be used to many SC domains; the model repository will contain the findings of different analytical techniques as well as stakeholder feedback. This repository, which will act as a knowledge repository or knowledge library, is where the gathered analytics and models are carefully documented, sorted, and classified. Information repositories assist stakeholders in making decisions in this way.

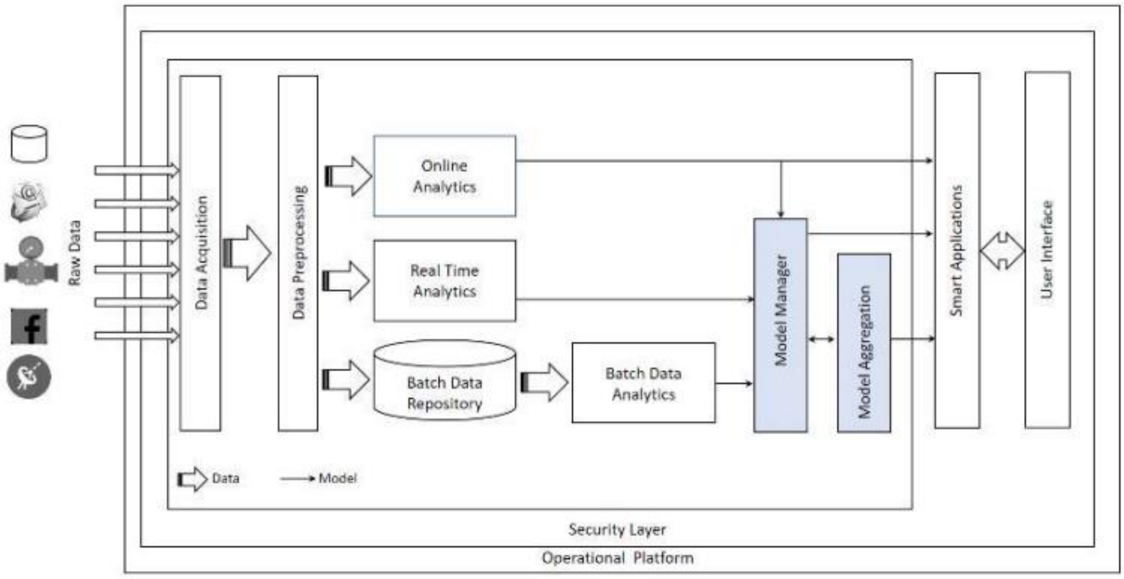


4. SCDAP—Prototypical Instantiation

We outline the SCDAP platform's physical implementation using Hadoop in this section. We first give a general overview of the SCDAP design concepts before getting into the specifics of the prototype instantiation. We will also go over the criticisms and remarks made in the studies that used SCDAP as a source. This stage tries to find any adjustments or errors that must be taken into account prior to the actual implementation of SCDAP.

5. Smart City Data Analytics Panel (SCDAP)

SCDAP is an end-to-end BDA framework for SCs that is not domain-specific. The layered design approach, standardised data gathering, access, real-time and historical data analytics, iterative and sequential development are the six design concepts that serve as the foundation of SCDAP's logical architecture. Intelligent Cities 2021, 4 DATA PROCESSING, EXTRACTED MODEL MANAGEMENT, AND AGGREGATION FOR PEER REVIEW 9



6. The Data Processing Layer

The elements required for data gathering, preparation, analysis, and display are included in the data processing layer. Rapid Miner is used in this layer. Rapid Miner is a data science software platform that offers an extensive and powerful library of operators—built-in processing elements for data analysis. Rapid Miner has a built-in library of operators, but it also adds a tons of third-party operators (including text mining, JSON parsers, deep learning, and Python scripting) to its library.

Rapid Miner is driven by Python and R scripting operators to add extra functionality, even though its ready-made operators speed up the development cycle of data analysis operations. One of the primary Rapid Miner operators that enable smooth connecting to Hadoop is Radoop. With more than 60 operators for sophisticated predictive modelling and data transformations that operate on a Hadoop cluster, Radoop expands the library of RapidMiner. The physical realisation mapping of the SCDAP prototype utilising CDH and RapidMiner is shown in Figure 6. Figure 6's SCDAP physical realisation components are identified by numbers that correspond to their respective conceptual design components.Finally, end users are shown the extracted findings using a Microsoft Power BI dashboard.

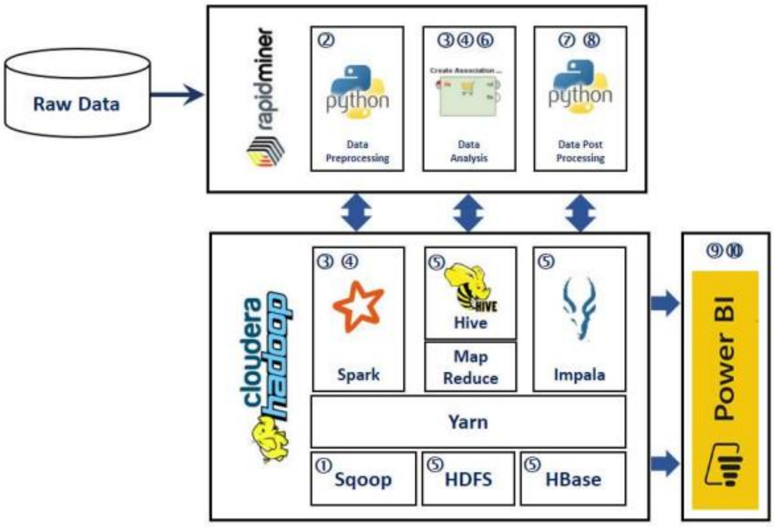


Figure 3: SCDAP prototype physical realization using CDH, Rapid Miner, and Power BI.

Data Preprocessing

Python scripts executed using the Rapid Miner Python operators are used to construct data preparation functions. In order to prepare the raw data for the analysis step, these features include tasks to eliminate abnormalities. Following are some examples of data preprocessing:

- Eliminating specialized characters from user names, company names, and user reviews
- Data integration: Bringing together different databases, data cubes, or files
- Data reduction: Dimensionality reduction, numerosity reduction, and compression
- Data cleansing: Filling in missing values and smoothing noisy data, identifying or removing outliers, and resolving inconsistencies

Data Analysis

The fundamental analytical elements to simplify and shorten development time are RapidMiner data analysis operators. We primarily concentrate on Meaning Cloud extension operators within the purview of this paper. Market-leading software for text analytics and semantic processing is called Meaning Cloud. It equips RapidMiner with a flexible and potent set of text analytics operators (such as topic extraction, sentiment analysis, and aspect-based sentiment analysis).

Data Post-processing and Model Management

The analysis phase's output data might be created in redundant or non-normalized formats that are unsuitable for model management and end-user display. Additional functions are needed to, if necessary, reorganise these output data before storing it for the subsequent steps. The metadata of the collected model may be managed by end users through intricate data entry and retrieval screens thanks to model management and aggregation features included in post-processing.

Conclusion

Big data is crucial for the development and administration of smart cities, to sum up. As urban regions become more populous and technologically advanced, the amount of data acquired from several sources, including sensors, devices, social media, and public services, is significantly increasing. Big data analytics may be used by smart cities to make use of this massive amount of data in order to get meaningful information that will enhance resource management, service delivery, and citizen quality of life.

Big data also promotes public participation and engagement. By leveraging data to give real-time information, residents of smart cities may make informed decisions about their daily lives, including transportation, energy consumption, and other aspects of it. Open data initiatives promote innovation and create new opportunities for economic growth by facilitating access to and analysis of regional data. By optimizing resource allocation, increasing service delivery, enhancing public safety, and promoting citizen engagement, big data has the potential to transform smart cities. It is essential to resolve privacy and ethical problems if big data benefits are to be fulfilled while preserving people's rights and interests. By striking a balance between data-driven decision making and moral data governance, smart cities may be able to leverage the promise of big data to create more sustainable, livable, and diversified urban settings.

References

1. Kitchin, R. (2015). The real-time city? Big data and smart urbanism. GeoJournal, 79(1), 1-14.
2. Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanović, N., &Meijers, E. (2007). Smart cities: Ranking of European medium-sized cities. Centre of Regional Science, Vienna UT.

3. Batty, M. (2013). Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274-279.
4. Chen, C., Zhang, C., Chen, L., &Choo, K. K. R. (2019). Big data analytics in smart cities: Concepts, methodologies, and applications. IEEE Communications Surveys & Tutorials, 21(1), 38-56.
5. Al Nuaimi, E., Al Neyadi, H., & Mohamed, N. (2015). Applications of big data to smart cities. Journal of Internet Services and Applications, 6(1), 1-15.
6. Zanella, A., Bui, N., Castellani, A., Vangelista, L., &Zorzi, M. (2014). Internet of things for smart cities. IEEE Internet of Things Journal, 1(1), 22-32.
7. Anthopoulos, L. G. (2015). Understanding the smart city domain: A literature review. In Proceedings of the 6th International Conference on Information, Intelligence, Systems and Applications (pp. 119-124). IEEE.
8. Yigitcanlar, T., &Kamruzzaman, M. (2018). Big data in smart cities.Cities, 81, 3-4.
9. F. Alassery, A. Alzahrani, A. Irshad Khan, A. Khan, M. Nadeem et al., "Quantitative evaluation of mental-health in type-2 diabetes patients through computational model," Intelligent Automation & Soft Computing, vol. 32, no.3, pp. 1701–1715, 2022.
10. Alyami, H.; Nadeem, M.; Alharbi, A.; Alosaimi, W.; Ansari, M.T.J.; Pandey, D.; Kumar, R.; Khan, R.A. The Evaluation of Software Security through Quantum Computing Techniques: A Durability Perspective. Appl. Sci. 2021, 11, 11784. https://doi.org/10.3390/app112411784
11. Suhel Ahmad Khan, Mohd Nadeem, Alka Agrawal, Raees Ahmad Khan, Rajeev Kumar, " Quantitative Analysis of Software Security through Fuzzy PROMETHEE-II Methodology: A Design Perspective", International Journal of Modern Education and Computer Science(IJMECS), Vol.13, No.6, pp. 30-41, 2021.DOI: 10.5815/ijmecs.2021.06.04
12. M. Ahmad, J. F. Al-Amri, A. F. Subahi, S. Khatri, A. Hussain Seh et al., "Healthcare device security assessment through computational methodology," Computer Systems Science and Engineering, vol. 41, no.2, pp. 811–828, 2022.
13. H. Alyami, M. Nadeem, W. Alosaimi, A. Alharbi, R. Kumar et al., "Analyzing the data of software security life-span: quantum computing era," Intelligent Automation & Soft Computing, vol. 31, no.2, pp. 707–716, 2022.
14. M. Nadeem, J. F. Al-Amri, A. F. Subahi, A. Hussain Seh, S. Ahmad Khan et al., "Multi-level hesitant fuzzy based model for usable-security assessment," Intelligent Automation & Soft Computing, vol. 31, no.1, pp. 61–82, 2022.

15. A. Alharbi, W. Alosaimi, H. Alyami, M. Nadeem, M. Faizan et al., "Managing software security risks through an integrated computational method," Intelligent Automation & Soft Computing, vol. 28, no.1, pp. 179–194, 2021.
16. A. Attaallah, S. Khatri, M. Nadeem, S. Anas Ansar, A. Kumar Pandey et al., "Prediction of covid-19 pandemic spread in kingdom of saudi arabia," Computer Systems Science and Engineering, vol. 37, no.3, pp. 313–329, 2021.
17. F. A. Alzahrani, M. Ahmad, M. Nadeem, R. Kumar and R. Ahmad Khan, "Integrity assessment of medical devices for improving hospital services," Computers, Materials & Continua, vol. 67, no.3, pp. 3619–3633, 2021.
18. A. Alharbi, M. Faizan, W. Alosaimi, H. Alyami, M. Nadeem et al., "A link analysis algorithm for identification of key hidden services," Computers, Materials & Continua, vol. 68, no.1, pp. 877–886, 2021.
19. Seh, A. H., Ahmad, M., Nadeem, M., Pandey, A. K., Agrawal, A., Kumar, R., & Khan, R. A. (2021). Usable-Security Assessment of Healthcare Software System Through Fuzzy ANP-TOPSIS Method. International Journal of System Dynamics Applications (IJSDA), 10(4), 1-24. http://doi.org/10.4018/IJSDA.304444
20. Alenezi, M., Nadeem, M., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Fuzzy multi criteria decision analysis method for assessing security design tactics for web applications. Int. J. Intell. Eng. Syst, 13(5), 181-196.