MACHINE LEARNING (AI) FOR IDENTIFYING SMART CITIES

L. Hammoumi1*, H. Rhinane1

1 Geoscience Laboratory, Ain Chock Faculty of Science, Km 8 El Jadida Road, BP:5366, Casablanca, Morocco - laazziza.hammoumi@gmail.com

KEY WORDS: Smart city indicators, Cities, Machine Learning, Artificial intelligence, Classification.

ABSTRACT:

Cities worldwide are attempting to be claimed as smart, but truly classifying as such remains a great challenge. This paper aims to use artificial intelligence AI to classify the smart city's performance as well as the factors linked to it. This is based on the perceptions of residents on issues related to structures and technology applications available in their cities. To achieve this goal, the study included 200 cities worldwide. For 147 cities we captured the perceptions of 120 residents in each city, by answering a survey of 39 questions evolving around two main Pillars: ‘Structures’ that refers to the existing infrastructure of the city and the ‘Technology’ pillar that describes the technological provisions and services available to the inhabitants. And each one is evaluated under five key areas: health and safety, mobility, activities, opportunities, and governance. The final score of the other 53 cities, was measured by using the data openly available on the internet. And this by means of different algorithms of machine learning such as Random Forest RF, Artificial Neural Network ANN, Support Vector Machine (SVM), and Gradient Boost (XGB). These algorithms have been compared and evaluated in order to select the best one. The tests showed that Random Forest RF alongside with Artificial Neural Network ANN, with the highest level of accuracy, are the best trained model. This study will enable other researches to use machine learning in the identification process of smart cities.

1. INTRODUCTION

The concept of smart cities has evolved over time, encompassing different aspects and definitions (Caragliu et al., 2011). Early discussions laid the foundations for understanding the multidimensional nature of smart cities. Harrison et al. (2010) presented a comprehensive view of smart cities, highlighting the integration of various technologies, data, and systems to enhance urban functions. Caragliu et al. (2011) further emphasized the role of information and communication technologies (ICT) in improving the efficiency and competitiveness of cities. As the concept matured, scholars recognized the importance of sustainability in smart city initiatives (Lombardi et al., 2012; Albino et al., 2015). The idea of a smart city gradually evolved into a holistic approach that seeks to address environmental, social, and economic challenges through innovative technologies and data-driven solutions (Harrison et al., 2010).

In the pursuit of identifying smart cities, researchers and practitioners have employed diverse methods and indicators. Sustainability has been a prominent aspect, with a focus on reducing resource consumption, promoting renewable energy, and improving environmental quality (Neirotti et al., 2014; Mora & Bolici, 2017). The notion of smartness, referring to the integration and utilization of advanced technologies, has also played a pivotal role in distinguishing smart cities (Batty, 2013; Angelidou, 2015). Moreover, performance indicators have been employed to evaluate the effectiveness and efficiency of smart city initiatives (Lai & Cole, 2022). These indicators encompass various dimensions such as governance, innovation, quality of life, and economic development (Huovila et al., 2019; Javed et al., 2022). By relying on such indicators, researchers and policymakers have attempted to quantify and compare the smartness of different cities (Albino et al., 2015).

However, these classical methods for identifying smart cities often rely on manual data collection and subjective assessments. This approach poses limitations in terms of scalability, objectivity, and accuracy (Sta, 2017). This is where the potential of machine learning (AI) techniques becomes evident, giving rise to the emergence of new methods and tools, such as Machine Learning (ML) techniques. Since then, machine learning has been widely applied across various domains, demonstrating its predictive capabilities and ability to extract meaningful patterns from vast amounts of data (Lim & Maglio, 2018). In the context of smart cities, machine learning holds promise for enhancing the identification and evaluation process (Khan et al., 2017).

Machine learning techniques can leverage big data analytics to provide valuable insights for smart city planning and management (Lim & Maglio, 2018; Hodorog et al., 2022). By analyzing large volumes of data generated by urban sensors, social media platforms, and other sources, machine learning algorithms can identify patterns, trends, and correlations that may not be readily apparent to human observers (Lim & Maglio, 2018; Hodorog et al., 2022). This data-driven approach can enable evidence-based decision-making, facilitate proactive interventions, and enhance the overall intelligence of cities (Kitchin, 2014). For instance, machine learning algorithms have been employed to predict disease outbreaks, optimize resource allocation, and improve the quality of healthcare services in the field of healthcare (Feng & Jiao, 2021). Similarly, in the domain of transportation, machine learning has been utilized to predict traffic congestion, optimize routing, and enhance mobility solutions (Shafiq et al., 2020).

While machine learning techniques have shown success in various domains, their application for identifying and defining smart cities has received limited attention (Khan et al., 2017). Most existing research has focused on utilizing machine learning for specific urban aspects, such as transportation, energy, or social media analysis (Khan et al., 2017; Bibri, 2018; Hodorog et al., 2022). By harnessing machine learning algorithms and techniques, researchers can develop models that learn from historical data and generate insights to identify and assess smart cities. For example, Kitchin (2014) discusses the potential of big data and smart urbanism in creating real-time cities, where machine learning algorithms can process massive
amounts of data and provide timely information for decision-making. Jamei et al. (2017) explored the role of virtual reality and machine learning in planning sustainable smart cities, highlighting the potential of these technologies in urban design and development. Additionally, Khan et al. (2017) proposed a framework for designing and planning smart cities based on big data analytics, which leverages machine learning algorithms to analyze and extract meaningful insights from large-scale urban data.

Machine learning holds the potential to predict future urban challenges and guide policy-making, with applications in predicting energy consumption patterns, optimizing resource allocation, and enhancing energy efficiency in buildings and infrastructure (Ullah et al., 2020; Duan et al., 2019), as well as supporting urban resilience and livability assessment for improved quality of life (Kutty et al., 2022).

Despite the numerous possibilities offered by machine learning, its successful integration into smart city identification and assessment requires addressing several challenges. One key challenge lies in the availability and quality of data. Machine learning models heavily rely on accurate and comprehensive data to generate reliable insights. Therefore, efforts should be made to ensure data accessibility, standardization, and quality control (Hashem et al., 2016). Additionally, ethical considerations, such as data privacy and algorithmic fairness, need to be taken into account to ensure responsible and inclusive deployment of machine learning techniques in smart cities (Allam & Dhunny, 2019).

While machine learning has been successfully applied in different domains, its use for identifying smart cities remains largely unexplored (Khan et al., 2017). This represents a significant gap in the current literature and presents an opportunity for future research to leverage machine learning as a tool to determine what a smart city truly encompasses. This research aims on harnessing the power of machine learning to enhance the identification and definition of smart cities, ultimately enabling more accurate, data-driven decision-making for urban development.

2. METHOD

Figure 1 depicts the proposed methodological approach, which can be summarized by the following steps: (1) Selection of smart cities, which will be the subject of our study or will constitute the experimental basis of our research; (2) Classification and processing of collected data from non-structured to semi-structured format; (3) Testing various machine learning algorithms; and (4) Results analysis and discussion. Each step will be detailed below.

2.1 Selection of smart cities

During our study, we relied upon the study articulated by Lai and Cole, in their research on “Measuring progress of smart cities: Indexing the smart city indices” published in Urban Governance (2022). The principal aim of this study is to critically examine the integrity and quality of existing smart city indices, subsequently discerning those indices that possess the requisite attributes for effective international comparative analysis.

As results from this study, we found that The Smart City Index (SCI), an ongoing initiative by the IMD World Competitiveness Center since 2019, has emerged as a pivotal tool within the domain of smart city assessment. Its annual evaluations, encompassing a diverse array of indicators across distinct categories, encapsulate a rigorous and comprehensive approach to assessing the multifaceted dimensions of smart urbanization. The SCI’s systematic framework, designed to gauge key aspects such as Health and Safety, Mobility, Activities, Opportunities, and Governance, lends credibility to its effectiveness as a robust evaluation mechanism. Furthermore, the SCI’s reliance on data gathered through citizen surveys enhances its utility, as it reflects the direct perceptions and sentiments of local residents. This participatory approach not only underscores its accuracy but also positions it as a reliable source for insights into citizens’ priorities and attitudes towards smart city development. As demonstrated by its consistent application and adaptability, the SCI stands as a dependable foundation for future research endeavors seeking to delve deeper into the complex landscape of smart urbanization.

The methodology employed in the IMD research encompasses a global assessment of 141 cities. The Smart City Index 2023 meticulously gauges residents’ perceptions concerning the urban infrastructure and technological applications available within their cities. It ranks 141 cities by soliciting the perceptions of 120 residents in each city. These perceptions are gathered within two pivotal pillars: Structures, focusing on existing infrastructure, and Technology, encapsulating available technological services. These pillars encompass key evaluation areas such as health and safety, mobility, activities, opportunities, and governance. Data, presented in tabular form, juxtaposes city scores against group benchmarks, facilitating comprehensive indicator comparisons.

To augment our research scope, we targeted 200 cities. So to enhance the comprehensiveness of our study, we expanded our focus to encompass an additional 53 cities. Among these, data for 47 cities were derived from openly accessible online sources. This research endeavor was conducted in-house, with the aim of thoroughly addressing the aforementioned indicators.

For the remaining 6 cities—Casablanca, Fez, Marrakech, Tangier, Dakhla, and Laayoune—we conducted an empirical survey. This survey was administered through a Google Form, comprising closed-ended questions meticulously aligned with the indicators aforementioned. The utilization of binary response options was chosen to facilitate both data transmission and subsequent analysis. The questionnaire was meticulously drafted in French, the prevailing language in the country, with the intention of reaching the widest possible audience. This approach was employed to maximize participation.

Figure 1. The proposed methodology.
2.2 Processing the collected data

The second section of the methodology expounds further into classifying and organizing the collected data, as described in section 1, from non-structured to semi-structured format. To do so, we transformed data into binary format that involves representing data using only two distinct values, typically 0 and 1. This representation aims to manipulate and transmit data. Each subtracted value from the survey that was lower than 50% was represented into the excel spreadsheet as the binary digit Zero (0), and subsequently every value greater than or equal to 50% was illustrated by bit one (1). The binary format enabled efficient data processing through simple logical operations.

The purpose of our study is to determine whether a given city qualifies as "smart." To accommodate this, we introduced an additional column labeled "target." Since we are working with a total of 39 indicators, we established a criterion where any city surpassing a strict cumulative score of 20 out of 39 would be recognized as "smart." This qualification is then denoted as a value of 1 in the "target" field for subsequent analysis.

2.3 Testing different algorithms of machine learning

The current study proposes a framework combining various machine learning models for the first time to thoroughly investigate the smartness of cities around the world, using a set of indicators.

The classification techniques, Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), support Vector Machine (SVM), and Gradient Boost (XGB) are used to predict the level of smartness, as categorical variables, based on the values of the indicators under each pillar: Structure and Technology. Multi-criteria performance assessment combines numerous heterogenous indicators across two main aspects in a standardized manner to a single synthetic score that explains the behavior of the phenomenon to be measured.

3. RESULTS

Initially, we evaluated the smartness of all 200 cities worldwide to answer the research question regarding the extent to which present-day smart cities engage in collaborative creation of infrastructure and technology within their development models. For each row, scores across each aspect under structure and technology pillars were added to forecast the smartness level of a city.

Different ML algorithms were trained on the data for the smart cities (a total of 7800 data points) to predict whether a city is smart or not, based on the values for the indicators under each aspect. Thus, the input vector for the assessment of the Structural aspect comprised 19 predictors. Similarly, for technological aspect, the indicators related to health & safety, mobility, activities, opportunities in work & school and governance (a total of 20 indicators) were used as predictors that determined the response variable, namely, the level of smartness. The dataset was split into train and test sets that comprised of 80% and 20% of the complete dataset, respectively.

To compare the classification models, we used the overall accuracy (ACC). Among the single models, ANN showed the highest performance in predicting the smartness of cities on the training dataset (97% accuracy), SVM and DT classifiers, on the other hand, showed a close performance of 95% and 92% of accuracy, respectively. Both, the ensemble models, RF and XGB showed a higher and equal accuracy compared to each other. They showed, similarly to ANN, the most accurate prediction on the test dataset, with 97.5% of accuracy each.

Building upon this analysis conducted on 200 cities under the two pivotal pillars—structural and technological—it was noteworthy that the study found a distinct pattern. Among the examined cities, a significant majority, accounting for 60%, were classified as 'smart', while the remaining 40% fell into the 'non-smart' category. This distribution implies a clear contrast in the developmental orientation and integration of these cities, forming the basis for understanding the defining factors that delineate 'smartness'.

Figure 2. Number Comparison of Smart and Non-Smart Cities.

In examining the structural pillars of smart city development across 200 cities worldwide (Fig. 3), a discernible trend emerges. Smart cities, identified as those showcasing a comprehensive and robust infrastructure alongside advanced technological provisions, exhibit distinctive characteristics within the structural dimension.

The data extracted from the study underscores a substantial correlation between certain indicators within the structural domain and the classification of cities as 'smart' or 'non-smart'. Notably, indicators related to fundamental necessities and community well-being prominently differentiate these classifications.

For cities categorized as 'smart' (indicated by 1), an evident trend surfaces, highlighting their strengths in critical infrastructure elements. Fig 3(a)- 3(b)- 3(e) shows factors such as adequate basic sanitation (SH 1), efficient recycling services (SH 2), robust medical service provision (SH 5), alongside efficient public transport (SM 8), abundant green spaces (SA 9), vibrant cultural activities (SA 10), accessible employment opportunities (SO 11), quality educational access (SO 12), lifelong learning prospects (SO 13), active job creation (SO 14), inclusive measures for minorities (SO 15), accessibility in local government decisions (SG 16) shown in Fig. 3(h) to Fig. 3(p) are notably higher in smart cities. Moreover, substantial feedback mechanisms on local government projects (SG 19) are evident, as illustrated in Fig. 3(s).

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-221-2024 | © Author(s) 2024. CC BY 4.0 License.
Figure 3. Structural Pillar Evaluation: Comparative Analysis of 19 Key Indicators Across Smart & Non-Smart Cities
Conversely, non-smart cities (designated as 0) tend to exhibit higher values in indicators pointing to persistent challenges. Factors in Fig. 3(c) - 3(d) respectively demonstrate public safety concerns (SH 3) and prevalent air pollution issues (SH 4) prevalent in non-smart cities. Furthermore, Fig. 3(f) - 3(g) portray housing affordability challenges (SH 6) and intensified traffic congestion (SM 7) prevalent among these cities. Moreover, pointing towards governance challenges, indicators SH17 and SH18 are depicted respectively in Fig. 3(q) - 3(r), showcasing concerns regarding corruption in governance (SG 17) and a lack of resident involvement in decision-making (SG 18).

The discernible disparities within these indicators imply a strong correlation between the quality of structural elements and the classification of cities as ‘smart’. It suggests that cities excelling in essential services, environmental quality, cultural vibrancy, educational opportunities, employment prospects, and community participation tend to align more closely with the definition of ‘smart cities’. This underscores the significance of robust infrastructure and community-focused services in shaping the smartness measure of cities on a global scale.

Meanwhile, in examining the technological pillar (Fig. 4) and its influence on the categorization of cities into ‘smart’ and ‘non-smart’, discernible patterns underscore the significance of specific indicators in differentiating between these classifications.

The data extracted from the study reveals a compelling trend among cities labeled as ‘smart’ (denoted by 1) regarding their technological infrastructure, showcased in Fig. 4. Notably, a significant percentage of these smart cities exhibit high values in specific indicators. Fig. 4(b)-(d) respectively represent online giving platforms (T2), efficient online city maintenance reporting (T3), and enhanced air pollution monitoring (T4). Furthermore, Fig. 5(f) portrays streamlined medical appointment arrangements (T6).

Furthering this pattern, from T10 to T14, Fig. 4(j)-(k)-(l)-(m)-(n) respectively showcase simplified public transport scheduling (T10), mobile-based traffic congestion updates (T11), easy online show and museum ticket purchases (T12), accessible online job listings (T13), and effective IT education in schools (T14). Additionally, Fig. 5(t) highlights the indicator T20, demonstrating the processing of identification documents online, effectively reducing waiting times. Remarkably, indicators T12, T13, and T16 stand out with nearly universal adoption among these ‘smart’ cities, depicted in Fig. 4(j)-(k) and 4(l) respectively.

Conversely, non-smart cities (represented by 0) tend to showcase higher values in indicators such as accessible online city maintenance reporting (T1) portrayed in Fig. 4(a); while Fig. 4(e) highlights streamlined public service-oriented Wi-Fi provision (T5). Furthermore, Fig. 4(g)-(h), and 4(i) respectively represent reduced congestion through car-sharing apps (T7), efficient parking location apps (T8), and enhanced bicycle sharing services (T9). Moreover, Fig. 5(o) showcases simplified public transport ticketing online (T15), whereas Fig. 4(q)-(r), and 4(s) respectively illustrate improved public access to city finances (T17), enhanced online voting systems (T18), and facilitated resident idea proposals (T19).

The distinction observed within these indicators underscores the pivotal role of specific technological aspects in categorizing cities as ‘smart’ or ‘non-smart’. It suggests that cities excelling in advanced technological integration, particularly in streamlined job access, online educational resources, and simplified identification procedures, are more aligned with the classification of ‘smart cities’. Conversely, cities focusing on accessible public services, efficient mobility solutions, and enhanced governance through technological means stand out among the ‘non-smart’ category.

4. DISCUSSION & CONCLUSION

This study introduced a novel assessment framework, combining multivariate data and diverse machine learning models to evaluate the smartness of 200 cities worldwide based on selected indicators. The proposed approach integrates the co-creation of structural and technological pillars within existing development models of smart cities. Analysis of multivariate data and city scores facilitates a comprehensive understanding of cities’ performance in structural and technological aspects, enabling ongoing performance monitoring for smart cities.

The utilization of different machine learning classifiers aimed to predict smartness levels for each city. Among these classifiers, the XGB and RF models emerged as the most accurate, as demonstrated by their high ACC parameter values. Ensemble modeling notably outperformed individual learning models, establishing its efficacy in predicting cities’ smartness levels.

The pivotal difference distinguishing ‘smart’ cities primarily lies within the realm of technological prowess. It’s the overwhelming prevalence of specific technological indicators that emphasizes the critical role of technological integration in shaping a city’s ‘smartness’. To earn the label of ‘smart,’ a city must prioritize and advance its technological infrastructure. This aspect serves as the distinctive feature and hallmark of comprehensive advancement, seamlessly integrating technology into daily life.

Figure 5. Comparative Analysis of Structural and Technological Pillars in Smart Cities.
Figure 4. Technological Pillar Evaluation: Comparative Analysis of 20 Key Indicators Across Smart & Non-Smart Cities.
Technological capabilities serve as the primary driver in distinguishing a city’s ‘smart’ status, outlining the imperative nature of technological advancement in modern urban development. Embracing innovation and integrating technology into urban infrastructure signify the evolution towards a smarter status. Such initiatives attract investments, expertise, and talent, fostering economic growth and defining the trajectory of smarter cities.

Understanding that no single city excels uniformly, it is crucial to recognize and learn from successful cities. Their models of urban development serve as blueprints for progress, attracting fresh ideas, jobs, and growth. Learning from these cities facilitates the development of livable and intelligent urban environments, essential for overall urban improvement. Ultimately, the success blueprint of each highly efficient city lies in its unique urban development model.

REFERENCES


Lim, C., & Maglio, P. P. (2018). Data-driven understanding of smart service systems through text mining. Service Science, 10(2).


