



Article

Predicting People's Concentration and Movements in a Smart City

Joao C. Ferreira ^{1,2,3} , Bruno Francisco ¹, Luis Elvas ^{1,2,3}, Miguel Nunes ¹ and Jose A. Afonso ^{4,*} 

¹ Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisboa, Portugal; jcafa@iscte-iul.pt (J.C.F.); bruno_alexandre_francisco@iscte-iul.pt (B.F.); luis.elvas@iscte-iul.pt (L.E.); miguel_bonacho@iscte-iul.pt (M.N.)

² Inov Inesc Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal

³ Department of Logistics, Molde University College, 6410 Molde, Norway

⁴ CMEMS-UMinho/LABELLS—Associate Laboratory, University of Minho, 4800-058 Guimarães, Portugal

* Correspondence: jose.afonso@dei.uminho.pt

Abstract: With the rapid growth of urbanization and the proliferation of mobile phone usage, smart city initiatives have gained momentum in leveraging data-driven insights to enhance urban planning and resource allocation. This paper proposes a novel approach for predicting people's concentration and movements within a smart city environment using mobile phone data provided by telecommunication operators. By harnessing the vast amount of anonymized and aggregated mobile phone data, we present a predictive framework that offers valuable insights into urban dynamics. The methodology involves collecting and processing location-based data obtained from telecommunication operators. Using machine learning techniques, including clustering and spatiotemporal analysis, we developed models to identify patterns in people's movements and concentration across various city regions. Our proposed approach considers factors such as time of day, day of the week, and special events to capture the intricate dynamics of urban activities. The predictive models presented in this paper demonstrate the ability to predict areas of high concentration of people, such as commercial districts during peak hours, as well as the people flow during the time. These insights have significant implications for urban planning, traffic management, and resource allocation. Our approach respects user privacy by working with aggregated and anonymized data, ensuring compliance with privacy regulations and ethical considerations. The proposed models were evaluated using real-world mobile phone data collected from a smart city environment in Lisbon, Portugal. The experimental results demonstrate the accuracy and effectiveness of our approach in predicting people's movements and concentration. This paper contributes to the growing field of smart city research by providing a data-driven solution for enhancing urban planning and resource allocation strategies. As cities continue to evolve, leveraging mobile phone data from telecommunication operators can lead to more efficient and sustainable urban environments.

Keywords: smart city; predictive modeling; concentration prediction; urban analytics and planning; data-driven decision making; neural networks



Citation: Ferreira, J.C.; Francisco, B.; Elvas, L.; Nunes, M.; Afonso, J.A. Predicting People's Concentration and Movements in a Smart City. *Electronics* **2024**, *13*, 96. <https://doi.org/10.3390/electronics13010096>

Academic Editor: Hung-Yu Chien

Received: 25 October 2023

Revised: 18 December 2023

Accepted: 21 December 2023

Published: 25 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, the concept of the smart city has gained considerable attention as urbanization accelerates and technology permeates every facet of modern life. A smart city employs data-driven strategies to enhance the quality of life for its residents by optimizing urban infrastructure, services, and resources. One of the key challenges in realizing the potential of smart cities lies in understanding and predicting the intricate dynamics of human movement and concentration within urban environments. This paper explores a pioneering approach to address this challenge, leveraging the ubiquity of mobile phones and the wealth of data they generate.

The rapid proliferation of mobile phones has transformed them into indispensable tools that accompany individuals in nearly every aspect of their daily routines. These

devices generate vast amounts of data, including the users' location, communication patterns, and interactions. Therefore, the telecommunication operators responsible for managing these networks possess an invaluable repository of aggregated and anonymized data that holds the potential to revolutionize urban planning, transportation management, and resource allocation.

Traditional methods of monitoring urban activity, such as static surveys or manual counting, are labor intensive, time consuming, and often fail to capture the dynamic nature of urban life. In contrast, mobile phone data offers a continuous and real-time stream of information that provides insights into how people move, gather, and distribute across the city landscape.

As a result, there is an increasing demand for efficient and low-cost urban studies, and some models have been created to help forecast the evolution of the urban expansion phenomena across time and place and to generate a more advanced understanding of them [1]. Urban growth models have risen in popularity over the past few decades, evolving from static monolithic models to autonomous dynamic models, despite the difficulty of effectively forecasting urban expansion through simulation methods [2]. Town planners and academics need to establish the driving elements that influence urban growth expansion in order to construct such models and forecast the future structure of a city [3]. To do this, they must identify data containing spatial details about the study location that could potentially have an impact. By identifying common patterns and quantifying their impact on urban expansion, statistical methods and algorithms can be used to ascertain the extent to which these factors have an impact.

This paper's motivation stems from the realization that harnessing mobile phone data from telecommunication operators can unlock a new dimension of understanding urban dynamics. The prediction of people's concentration and movements within a smart city environment enables a more efficient allocation of resources, better traffic management, improved emergency response, and enhanced urban design. Furthermore, such insights can empower policymakers to make informed decisions that lead to sustainable development, reduced congestion, and enhanced quality of life for city residents.

The ethical and privacy considerations inherent in working with personal data are of paramount importance. Therefore, this paper emphasizes the use of aggregated and anonymized data to respect user privacy and comply with the relevant regulations. Through this approach, we aim to balance the potential benefits of data-driven urban insights with the ethical imperative of protecting individual privacy.

In this context, our paper introduces a novel predictive framework that utilizes mobile phone data from telecommunication operators to predict people's movements and concentration patterns in a smart city environment using neural networks (NNs), given the massive amount of data available. Through bridging the gap between cutting-edge technologies and urban planning, we contribute to the growing body of research that seeks to create smarter, more responsive, and more sustainable urban environments.

The purpose of this research is to construct a people's concentration model that takes advantage of the potential of machine learning. As one of the main contributions of this work, we offer a way for gathering, handling, organizing, and, finally, converting primary open geographic data into tabular information suitable for machine learning algorithms and techniques. We also offer a multi-layered neural network and detail the many parameters used in the model in addition to the techniques and functions employed. In addition, we offer a collection of Python scripts that can be used to carry out all the mentioned methods and approaches of this study as an open-source GitHub repository.

Our predictive framework employs advanced machine learning techniques, including clustering and spatiotemporal analysis, to unveil patterns in human movement and concentration across diverse city regions. Rigorous evaluation with real-world mobile phone data demonstrates the accuracy and effectiveness of our models in anticipating high-concentration areas and predicting people's movements across various times and circumstances.

The implications of our research are substantial. Urban planners, policymakers, and city officials can leverage our approach's insights to make informed decisions on transportation infrastructure, emergency-response planning, crowd management, and urban design. Through embracing predictive capabilities, cities can strive for greater efficiency, sustainability, and livability.

Emphasizing individual privacy and ethical considerations, we underscore the use of anonymized and aggregated data to ensure the benefits of our predictive models without compromising privacy rights or violating regulatory norms.

While presenting a novel predictive framework using neural networks, our paper also sets the stage for future research. Integration of additional data sources such as weather patterns, social media activity, and public events could enhance prediction accuracy. Further refinements to accommodate evolving urban dynamics and exploring real-time predictions offer potential avenues for more actionable insights in smart city initiatives.

The predictive analytics process has been integrated into an interactive dashboard, providing experts with the capability to navigate seamlessly through both temporal and spatial dimensions. Utilizing the output predictions, we generated spatial-temporal heat maps, enabling experts to assess and monitor people concentration effectively.

The subsequent sections of this paper are structured as follows: Section 2 reviews the related literature concerning the prediction of people's concentration and movements within a smart city. In Section 3, we detail the methodology employed in this work, focusing on the identification of mobility patterns through the CRISP-DM framework [4], used by others in similar problems [5,6]. Section 4 is dedicated to the experimental results and the evaluation of the prediction process. Moving forward, Section 5 provides insights into the developed dashboard, encompassing both past data and the accomplished predictions. Finally, Section 6 offers conclusions drawn from our findings and provides suggestions for future research endeavors.

2. Related Work

The emergence of smart cities has introduced novel opportunities for leveraging data-driven approaches to enhance urban planning, resource allocation, and overall quality of life. With the proliferation of mobile phones, these devices have become an integral part of modern urban living, generating a wealth of data that can provide valuable insights into human behavior and movement patterns. This section provides a literature review of state-of-the-art research that explores works related to the prediction of people's concentration and movements within a smart city environment using mobile phone data provided by telecommunication operators.

In [7], Zhang et al. proposed a machine-learning approach to predict human mobility patterns in a smart city. The authors leverage various data sources, including mobile device data, social media data, and transportation data, to train predictive models. The study demonstrates the effectiveness of the approach in accurately forecasting people's concentration and movements in different urban areas.

In [8], Ma et al. reviewed and synthesized existing studies on individual mobility prediction in transport (data/problem/methodology/applications), identifying remaining research needs and discussing methodological considerations and potential future transport applications.

In [9], Li et al. proposed a method for predicting urban human mobility patterns using large-scale taxi traces. They leverage the wealth of data collected from taxis, including GPS traces, to analyze and understand human mobility in urban areas. The study focuses on the prediction of future human movement patterns based on historical taxi trajectory data. The authors employ machine learning and data mining techniques to process and analyze the taxi trace data, developing prediction models that consider various factors, such as time of day, day of the week, and geographical features. Through training these models with historical taxi traces, they aim to predict the future movement patterns of individuals within the city. Furthermore, the paper discusses the applications of the proposed prediction model

in various urban planning and transportation-management scenarios. The ability to predict human mobility patterns can aid in optimizing transportation systems, urban infrastructure planning, and resource allocation. The authors highlight the potential benefits of using large-scale taxi trace data for understanding and predicting urban human mobility and its impact on the city's overall functionality. Overall, the paper presents a methodological approach to predict urban human mobility using large-scale taxi traces and highlights the potential applications of such predictions in urban planning and transportation management.

Zhou et al. [10] presented a deep-learning-based approach for predicting human mobility patterns in smart cities. The authors propose a long short-term memory (LSTM) network that takes into account spatiotemporal features extracted from mobile device data and transportation data. The experimental results demonstrate the superior performance of the LSTM network machine learning algorithms in accurately forecasting people's concentration and movements.

Yao et al. [11] focused on predicting urban crowd density using mobile sensing data. The authors propose a prediction model that combines mobile device data, such as cellular network signals and Wi-Fi data, with location information. The model employs a regression-based approach to forecast crowd density in different areas of a smart city. The experimental results show promising accuracy in predicting crowd concentration.

The utilization of mobile phone data for urban analysis has gained traction in recent years. Researchers have demonstrated the potential of call detail records (CDRs) and location data to uncover valuable insights into population movement, social interactions, and transportation patterns. In [12], Phithakkitnukoon et al. used mobile phone data to create dynamic mobility maps, revealing spatiotemporal patterns in urban movement. This foundational work established the feasibility of using mobile phone data for urban analysis.

The study of human mobility patterns has also attracted considerable attention due to its implications for urban planning and transportation management. In [13], Toole et al. analyzed CDRs to identify regular mobility patterns. They found that, despite the diversity of individual movements, collective patterns exhibit high predictability. Understanding these patterns is crucial for predicting areas of high concentration and movement within a smart city.

Predictive modeling techniques have been applied to mobile phone data to anticipate urban dynamics. In [14], Wang et al. introduced a model that leverages historical data to predict the next location of individuals, contributing to the understanding of mobility patterns. In [15], Kothari et al. proposed a trajectory-based approach that captures individual movement history to predict future trajectories. These studies demonstrate the potential of predictive models in forecasting people's movements within urban environments.

Smart city initiatives focus on enhancing urban planning through technology and data-driven strategies. In [16], Giffinger et al. provided a framework for defining smart cities, emphasizing the importance of innovation, technology, and sustainability. Predicting people's concentration and movement aligns with the smart city vision through providing actionable insights for optimized land use, transportation planning, and resource allocation.

Utilizing personal mobile phone data raises concerns about individual privacy. Researchers have addressed these concerns through advocating for anonymization techniques and aggregate-level analysis. In [17], De Montjoye et al. demonstrated that even when anonymized, human mobility data can still be re-identified, highlighting the need for careful data handling to ensure privacy protection.

The integration of mobile phone data into practical urban applications has shown promising results. In [18], Hong et al. utilized mobile phone data to estimate traffic congestion, providing real-time traffic information for urban commuters, whereas in [19], Calabrese et al. introduced an approach to predict urban crowd dynamics, aiding in the planning of public events and large gatherings. The works related to urban mobility are resumed in Table 1.

Table 1. Summary of the related work implying urban mobility and urban information.

Ref.	Authors	Main Topics	Methodology	Key Findings and Applications
[7]	Zhang et al.	Human Mobility Prediction	Machine-learning approach using mobile device data, social media data, and transportation data for human mobility prediction in smart cities	Effective forecasting of people's concentration and movements in urban areas, demonstrating the utility of diverse data sources in predictive models.
[8]	Ma et al.	Mobility Prediction Review	Review and synthesis of studies on individual mobility prediction in transport, addressing data, problems, methodology, and applications	Identified research gaps, discussed methodological considerations, and explored potential future transport applications in individual mobility prediction.
[9]	Li et al.	Urban Mobility Patterns	Prediction of urban human mobility patterns using large-scale taxi traces with machine learning and data mining techniques	Focus on predicting future human movement patterns based on historical taxi trajectory data, highlighting applications in urban planning and transportation management.
[10]	Zhou et al.	Deep Learning for Mobility	Deep learning-based approach using LSTM network for predicting human mobility patterns in smart cities	Demonstrated superior performance of LSTM network in accurately forecasting people's concentration and movements, leveraging spatiotemporal features from mobile device and transportation data.
[11]	Yao et al.	Urban Crowd Density	Prediction of urban crowd density using mobile sensing data and regression-based approach	Proposed a model combining mobile device data and location information for accurate forecasting of crowd density in smart cities.
[12]	Phithakkitnukoon et al.	Mobile Phone Data Analysis	Utilized mobile phone data to create dynamic mobility maps, revealing spatiotemporal patterns in urban movement	Demonstrated the feasibility of using call detail records (CDRs) and location data for urban analysis, providing valuable insights into population movement, social interactions, and transportation patterns.
[13]	Toole et al.	Regular Mobility Patterns	Analyzed CDRs to identify regular mobility patterns in smart cities	Despite individual movement diversity, identified high predictability in collective patterns, crucial for predicting areas of high concentration and movement within a smart city.
[14]	Wang et al.	Next Location Prediction	Introduced a model leveraging historical mobile phone data to predict individuals' next locations	Contributed to understanding mobility patterns within urban environments, highlighting the potential of predictive models in forecasting people's movements.

Table 1. Cont.

Ref.	Authors	Main Topics	Methodology	Key Findings and Applications
[15]	Kothari et al.	Trajectory-Based Prediction	Proposed a trajectory-based approach capturing individual movement history to predict future trajectories	Demonstrated the potential of predictive models in anticipating people's movements within urban environments.
[16]	Giffinger et al.	Smart Cities Framework	Provided a framework for defining smart cities, emphasizing innovation, technology, and sustainability	Emphasized the alignment of predicting people's concentration and movements with the smart city vision, providing actionable insights for optimized land use, transportation planning, and resource allocation.
[17]	De Montjoye et al.	Privacy in Mobile Data	Addressed privacy concerns in utilizing mobile phone data, advocated for anonymization techniques and aggregate-level analysis	Demonstrated that even when anonymized, human mobility data can still be re-identified, highlighting the need for careful data handling to ensure privacy protection.
[18]	Hong et al.	Traffic Congestion	Utilized mobile phone data to estimate traffic congestion, providing real-time traffic information for urban commuters	Contributed to practical urban applications, offering real-time traffic information for commuters using mobile phone data.
[19]	Calabrese et al.	Urban Crowd Dynamics	Introduced an approach to predict urban crowd dynamics for planning public events and large gatherings	Demonstrated the integration of mobile phone data into urban applications, aiding in the planning of public events and management of large gatherings.
[20]	Calabrese et al.	Urban Growth Effects	Described negative environmental effects of rapid urbanization and the need for responsible land use planning	Highlighted the environmental and socioeconomic effects of urban expansion, emphasizing the importance of responsible land use planning.
[21]	Nikitas et al.	Land Use Projection Models	Discussed models for projecting changes in land use, including cellular automata (CA), agent-based models (ABM), and machine learning models	Introduced spatiotemporal models for projecting changes in land use, categorizing models into CA, ABM, and machine learning, with a focus on their application in examining the effects of urban land growth.

In summary, this literature review highlights the growing body of research focused on the potential of using mobile phone data from telecommunication operators to predict people's concentration and movements within a smart city. The convergence of urban planning, data science, and technology offers opportunities to create more efficient and sustainable urban environments. However, ethical considerations and privacy safeguards must be a central focus when working with personal data to ensure that the benefits of predictive models align with responsible-data-usage practices. This paper aims to contribute to this evolving field by presenting a predictive framework that addresses these challenges and advances the understanding of urban dynamics.

A method called land-use planning makes sure that those who live there use the land responsibly. However, significant growth brought on by rapid urbanization causes negative environmental effects such as air pollution, overuse of natural resources, and traffic congestion [20]. The negative socioeconomic effects of urban expansion can include poverty, unemployment, and restricted access to social services [21]. In order to enable town planners to examine all of these effects using fictitious scenarios of urban land growth, a number of models for projecting changes in land use have been developed [22].

These models, which can be categorized as cellular automata (CA), agent-based models (ABM), and machine learning models, may be referred to as spatiotemporal models [1]. Nevertheless, the idea of machine learning encompasses a variety of methodologies and sub-techniques, including logistic regression (LR), artificial neural networks (ANN), linear regression (LN), and decision trees (DT). Because of its simplicity, versatility, excellent performance, and in particular due to its capacity to capture the spatial and temporal elements of urban activities, cellular automata may be regarded as the most well-studied instance. Simply put, CA is a collection of ordinary cells, each of which is constantly holding a value from the set of permitted states.

The majority of the time, this is a binary value (build/not-build), and its state varies in response to the states of the cells around it. Every cell in the dataset experiences the application of a set of transition rules at each cycle, generating the state value for the following cycle. Several researchers [23–25] have employed CAs to anticipate urban growth ranging from tiny to large cities in different geographic locations of the world.

ABM, a dynamic computational model for simulating the activities and behaviors of autonomous agents, is similar to the CA model. According to a set of predetermined rules, agents are able to independently evaluate their current situation and take action [26]. As a result, ABMs may be thought of as more potent models than CA as agents are allowed to move around and interact with one another and with their surroundings throughout time. Furthermore, because they have the capacity to store a variety of diverse driving forces for social and environmental phenomena, agents may be extremely complex. Numerous urban studies have also utilized ABMs [27,28].

Due to its strength in terms of performance and accuracy while working with enormous datasets, the machine learning model has become one of the modern methodologies that have gained prominence over the previous ten years [29]. Up to now, spatial modelers have been able to create these models by either employing model platforms like GAMA (GIS Agent-based Modelling Architecture) or through creating model plugins for for-profit and free desktop GIS (Geographic Information System) apps [30]. Today, spatial modelers may create models with higher performance and accuracy with just a few lines of code and libraries like Keras [31] or Gdal [32]. Additionally, there are many different statistical techniques and algorithms available to modelers. However, given that the use of machine learning in urban studies is still in its infancy, it is anticipated that it will develop further and, as a result, enhance the accuracy and performance of urban growth models.

3. CRISP-DM Implementation towards Identifying Mobility Patterns

3.1. Business Understanding

Besides forecasting the number of people in a given area at a given period with reasonable accuracy, it is desirable to identify from which neighboring areas people tend to

enter and through which ones they tend to exit. This information allows the city council to make more informed decisions ahead of time and intervene in situations where it is relevant to know more than just the number of people in a particular area. One way to achieve this is to identify the mobility patterns in the areas of interest (nightlife areas in the context of this study), which allows for an understanding of how people move in a geographical area.

3.2. Data Understanding

The primary dataset utilized in this project was initially compiled by the telecommunications company in Lisbon and was subsequently provided by the Lisbon City Council. This dataset encompasses information related to the mobile phone numbers entering, remaining, and exiting 3743 grid cells measuring 200 by 200 m in Lisbon, spanning from 15 September 2021, to 31 December 2022. Comprising 24 variables, the dataset underwent further processing by the Lisbon City Council to accurately reflect the count of mobile phones across all telecommunications companies within each grid.

3.3. Data Preparation

The pre-processing phase was expedited due to the well-organized nature of the provided data. During the analysis to identify missing, duplicate, and misformatted data, it was observed that dates without records were notably absent across all variables. Additionally, variables D1 and E6 contained NaN values and were deemed irrelevant to our project objectives. Consequently, these variables were excluded, resulting in a pristine dataset devoid of any NaN values.

However, it is noteworthy that, during practical considerations, dates without records were identified but not incorporated into the dataset, revealing underlying data discrepancies that were subsequently rectified in later stages of our analysis.

The need to construct an uninterrupted time series prompted a thorough investigation into potential flaws within the provided dataset. The analysis unfolded in two stages: an initial phase identifying issues in all grids persisting over 24 h and a subsequent, more detailed, examination pinpointing flaws in 5 min intervals for each grid individually.

During the first stage, conducted for October 2021, several problems were uncovered, including misformatted and duplicated data. A complete data breakdown between February and March was identified, with remaining months showing varying flaws that necessitated further scrutiny.

The second stage of analysis, spanning January to December 2022, unveiled all flaws, distinguishing between total and partial disruptions. Total flaws indicated failures across all grids, with one notable interruption occurring from 8 February 2022, at 12:05 a.m., to 18 March 2022, at 10:40 a.m., lasting for 38 days, 10 h, and 35 min—a likely consequence of the Vodafone cyberattack. Given the substantial disruptions, the analysis focused on the second part of the data, spanning from 18 March 2022, to 31 December 2022.

This phase also highlighted additional significant failures, such as the two-day lapses from 19 to 21 November 2021, and 28 to 30 June 2022. Noteworthy single-day failures occurred on 30 October 2022 (1 day, 7 h, and 45 min), and 19 November 2022 (1 day, 10 h, and 15 min).

Detailed records of partial data failures, including duration and the number of affected grids on respective dates, were compiled. Notable instances on 18–19 January 2022 (25 partial failures), 29–30 September 2022 (17 combined), and 10 November 2022 (20 failures), highlighted specific dates with multiple failures.

All the information was compiled into 3743 square grids with 200×200 m. The information is gathered at intervals of 5 min. Due to privacy restrictions, if a particular grid has fewer than 10 users during a 5 min period, this information is omitted. After being collected, this information is made available on the dashboard platform for around 45 min. This means that the time between data collection and availability can only be delayed by

a maximum of one hour. It is crucial to note, though, that for the purposes of the current study, we used just a snapshot of the data instead of utilizing the online data stream.

Another dataset that provides information about each of the 3743 grids was also used in addition to the main dataset that comprises the data provided by the city council through the agreement reached with a mobile operator. This additional dataset includes the coordinates of each grid's centroid, the parish or parishes in which the grid is inserted, its name, geometry and WKT (Well-Known Text), which enable us to geo-reference the data from the main dataset. With this knowledge, the events may be added to the space using the Grid_ID key, which then launches our analyses.

Table 2 provides a list of all variables (features) contained in the main dataset as well as a description of the respective data.

Table 2. Mobile operator dataset variables.

ID	Variable Name	Variable Description	Variable Type
0	Grid_ID	Number of the grid. There are 3743 squares of 200 by 200 m to cover the metropolitan area of Lisbon	Nominal
1	Datetime	Time and date of occurrence	Datetime
2	C1	Number of distinct terminals counted on each grid cell during the 5 min period, measured every 5 min	Metric
3	C2	Number of distinct terminals in roaming counted on each grid cell during the 5 min period, measured every 5 min	Metric
4	C3	Number of distinct terminals that remained in the grid cell counted at the end of each 5 min period	Metric
5	C4	Number of distinct terminals in roaming that remained in the grid cell counted at the end of each 5 min period	Metric
6	C5	Number of distinct terminals entering the grid	Metric
7	C6	Number of terminals leaving the grid. These are the distinct terminals that left the grid. The calculation is made using the previous 5 min interval as reference, also considering the crossings of the grid in the same interval	Metric
8	C7	Number of entries of distinct terminals, in roaming, in the grid	Metric
9	C8	Number of exits of distinct terminals, in roaming, in the grid	Metric
10	C9	Total number of distinct terminals with active data connection in the grid cell. Measurement every 5 min	Metric
11	C10	Total number of distinct terminals, in roaming, with active data connection in the grid cell. Measurement every 5 min	Metric
12	C11	Number of voice calls originating from the grid cell	Metric
13	C12	Entering the city: Number of devices that for 5 min enter the 11 street sections considered for analysis	Metric
14	C13	Entering the city: Number of devices that for 5 min enter the 11 street sections considered for analysis	Metric
15	D1	Top-10 origin countries of the devices in roaming	Metric
16	E1	Number of voice calls that ended in the grid within the 5 min	Metric
17	E2	Average download speed per grid within the 5 min	Metric
18	E3	Average load speed per grid within the 5 min	Metric
19	E4	Peak download speed on the grid within the 5 min	Metric

Table 2. Cont.

ID	Variable Name	Variable Description	Variable Type
20	E5	Peak upload speed on the grid within the 5 min	Metric
21	E6	Top 10 apps used on the grid within the 5 min	Metric
22	E7	Lowest permanence period on the grid within the 5 min	Metric
23	E8	Average permanence on the grid within the 5 min	Metric
24	E9	Maximum permanence period on the grid within the 5 min	Metric
25	E10	Number of devices sharing the Internet connection in the grid within the 5 min	Metric

After gathering the necessary information for our study, we carefully reviewed it and analyzed each variable to determine its potential and how we could improve the added value of this study. As said before, our main goal is to comprehend how tourists move around the city. In order to do this, we used the C1, C2, C3, C4, and C5 features from Table 1, regarding people's movement in Lisbon (both roaming and non-roaming), which was based on the data generated by mobile phones.

3.4. Statistical Models

Frias-Martinez et al. [33] estimated the daily mobility patterns in a city based on data collected from mobile phones. Their aim was similar to the one in this section, but the difference being their focus on daily patterns in the city. However, the researchers used mobile phone records, where they had the geo-coordinates from the callers and the receivers, and their methodology consisted in following each individual mobile phone. The results were positive, but given that the dataset provided by the city council does not allow the tracking of each individual terminal for privacy reasons, it was not possible to replicate such a method in this paper. Other similar works, such as Li et al. [6], also used data collected from mobile phones where each user was individually identified.

A possible approach to this identification is directly related to the models in the STARMA (Seasonal, Trend, AutoRegressive, Moving Average) family. The STARMA family of models is an extension of the traditional ARMA (AutoRegressive Moving Average) time series models. These models are used for modeling and forecasting time series data that exhibit seasonal patterns, trends, and autoregressive and moving average components. STARMA models incorporate both autoregressive and moving average terms, as well as seasonal and trend components, to capture complex patterns in time series data.

For simplicity, moving average and autoregressive components are frequently abbreviated as "STARMA". To better understand its underlying patterns and produce more precise projections, this technique separates and analyzes time series data into these four essential parts.

1. The seasonal component, such as daily, monthly, or yearly patterns, is a regular, recurring pattern or fluctuation in the time series data that occurs at intervals. It aids in the identification of seasonality, which can be crucial in a variety of applications like retail sales forecasting, where seasonal spikes may frequently occur.
2. Long-term movement or direction from the time series data is captured by the trend component, which aids in determining if the data is typically rising, falling, or remaining stable over time. It is for understanding broad growth or fall trends in data, such as stock prices or population statistics.
3. The autoregressive component (AR) simulates the relationship between the time series' present value and its previous values. It quantifies the relationship between the current data point and its own lag values, being very useful for identifying temporary dependencies and data.

- The moving average (MA) component considers the data's short-term volatility or noise by accounting for prior forecast errors. By removing random fluctuations from the time series, it makes it simpler to spot underlying patterns.

Analysts and researchers can better understand a time series' behavior through breaking it down into these four parts. They can then use this information to analyze the effects of seasonality and patterns on the data, reduce noise, and generate more precise forecasts and predictions.

Through the creation of a precise spatial weights matrix that effectively captures the interconnections between neighboring areas, we can discern the mobility patterns within the nightlife zones. A substantial weight assigned to a particular neighbor signifies a significant flow of people between these areas. Accordingly, we modified the STARMA model and explored various approaches for constructing an accurate spatial weights matrix, which encompassed:

- Comparing the number of exits with the neighbors' number of entries.
- Comparing the time series of an area and its neighbors.

All these approaches have the same limitation: it is impossible to accurately determine where people tend to go when they leave an area. Let us consider Figure 1, where a potential issue is illustrated.



Figure 1. Example of entries and exits for a grid with its neighbors.

If the method of comparing the number of entries and exits was chosen, one would look at the example in Figure 1 and conclude that the people who left this specific area tended to go to neighbor A. In this specific instance, the observed behavior did not conform to the anticipated pattern. Those departing the area chose to visit every neighboring location except for A, and neighboring entries predominantly originated from other sources. If we had assumed that the high volume of exits from Figure 1 held significance, it would have led to erroneous conclusions. Consequently, the project will not yield insights into the mobility patterns within the nightlife areas.

However, we consider the Wasserstein distance [34] to estimate these patterns in future work. Balzotti et al. [29] show that this approach can produce positive results.

Mobile operators continuously gather information about how customers and users utilize the service, whether for technical, monetary, or even legal reasons. In this regard, the network must constantly gather various data and metrics regardless of whether the customer is using 2G, 3G, 4G, or 5G, in order, for example, to allow the continuation of a phone call or a mobile data session. It should be highlighted that when we refer to "data from a mobile operator", we do not mean the mobile data service, but rather detailed information about all the signaling exchanged between the user's mobile device and the network providing the service. Although an operator may have multiple networks,

depending on the technology used, there is always a need for a base station that is in a specific location and serves customers in a specific geographic area. This base station interacts with the mobile phone while also interacting with the network core, and through this interface, it gathers events like a voice call, an Internet session or a written message, in addition to event data, such as a network attach or detach, handover, location updates, etc. The analysis and the dashboard created for this paper were based on the signal exchanged in the Lisbon city through a 12-month period considering roamers and national users.

It was also possible to correlate each event with a specific location using a metadata file, which was used to supplement and augment the data set provided by the mobile operator. The information provided by the Lisbon city council was generated using data from each user's mobile phone device and cellular network as part of a contract with a mobile operator. For legal and privacy considerations, the data in the dataset has been appropriately anonymized. This makes it impossible to perform any type of precise analysis on a specific user. Analysis can only be conducted using the aggregated data, and there is not any key in the data linking a specific person to an event.

3.5. Neural Networks

Regarding the machine learning methods, since we have big data (each month is around 2 GB), neural networks were applied. Within this type, the LSTM networks are adequate because we need to handle time and space. At an initial phase, we compared time series and random forest, but LSTM results were superior. Due to its ability to handle ample time lags [35], it is a very popular deep-learning-based method for time series forecasting.

In this project, Bidirectional LSTM (BiLSTM), a variation of LSTM was used to make the forecasts. This particular kind of LSTM distinguishes itself from the others because of its ability to have the input layer fed both forwards and backwards during processing [36]. With a series of experiments, it was determined that BiLSTM gave better results than conventional LSTM in the context of this particular dataset.

The dependent variable in this model is the number of different terminals in the area (C1). The independent variables, besides the number of different terminals in the past, are the difference between the number of entries and exits, and a dummy variable that has a value of 1 when the period of the observation corresponds to a period of known nightlife (from 11 p.m. on Friday to 5 a.m. on Saturday, and from 11 p.m. on Saturday to 5 p.m. of Sunday), and 0 otherwise.

The model used the past four occurrences to make the forecasts (equivalent to 20 min). The activation function was ReLU (Rectified Linear Unit), and the number of epochs (passes through the training dataset) was limited to 80. Additionally, if no improvements were detected after 20 epochs, the model was instructed to stop its fitting stage.

Given that LSTM only considers the temporal component of the dataset, each one of the 161 identified nightlife areas was fitted with its own network. Their time series began on April 1st and ended on December 31st, and were divided into three sets: training set (75%), validation set (15%), and test set (10%). Due to the data's variety, the forecasts quality was mixed.

When fitting the model, TensorFlow used the root mean squared error (RMSE) of the validation set to assess if a model was better than the past ones to avoid overfitting.

4. Experimental Results

During the model fitting process, TensorFlow employed the Root Mean Squared Error (RMSE) of the validation set as a criterion for assessing model performance, ensuring improvements over previous iterations to guard against overfitting. This evaluation metric was further extended to the training and test sets, with the distribution of values for the latter visualized in Figure 2 (left image). However, due to the considerable variability in maximum and mean values across distinct areas, the RMSE metric was deemed unsuitable for making meaningful comparisons. Consequently, an additional metric, the Mean Absolute Percentage Error (MAPE), was introduced and evaluated, as illustrated in Figure 2

(right image). This strategic inclusion of MAPE enhances the robustness of our assessments, providing a more nuanced perspective on forecast accuracy that accounts for the diverse characteristics of each area. The majority of forecasts exhibited a commendable accuracy level, with the Mean Absolute Percentage Error (MAPE) typically below 20, signifying reliable predictions. Nevertheless, it is imperative to scrutinize the LSTM’s ability to effectively anticipate notable spikes in the number of terminals. This critical assessment is exemplified in Figure 3 (left image), showcasing the network’s performance in an area characterized by a high MAPE of 68.2%, and in Figure 3 (right image), illustrating the network’s efficiency in an area with a low MAPE of 10.3%. These contrasting scenarios offer insights into the model’s performance across varying levels of prediction challenges, emphasizing the need for a nuanced evaluation beyond average accuracy metrics.

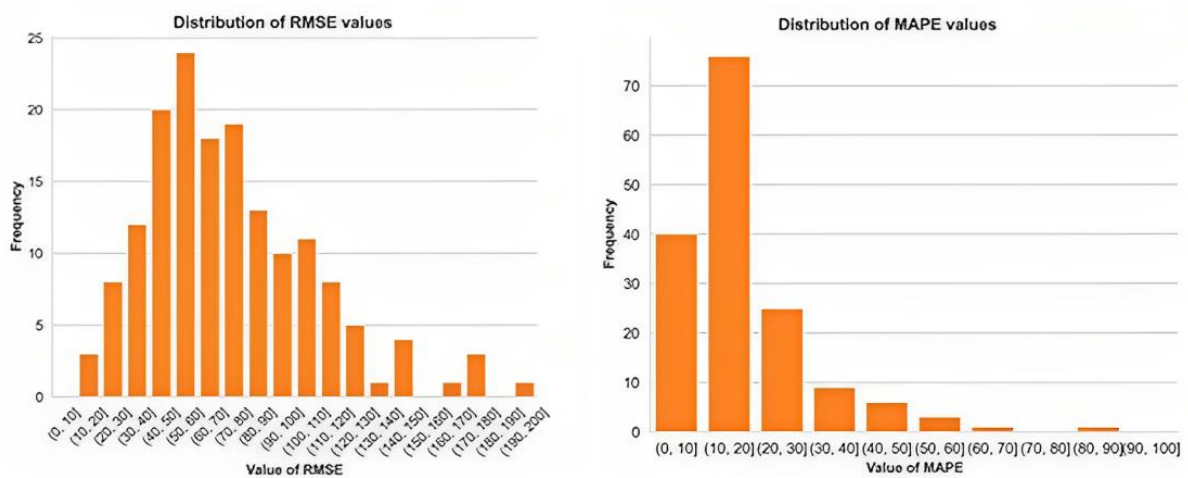


Figure 2. Distribution of RMSE (left image) and MAPE (right image) values.

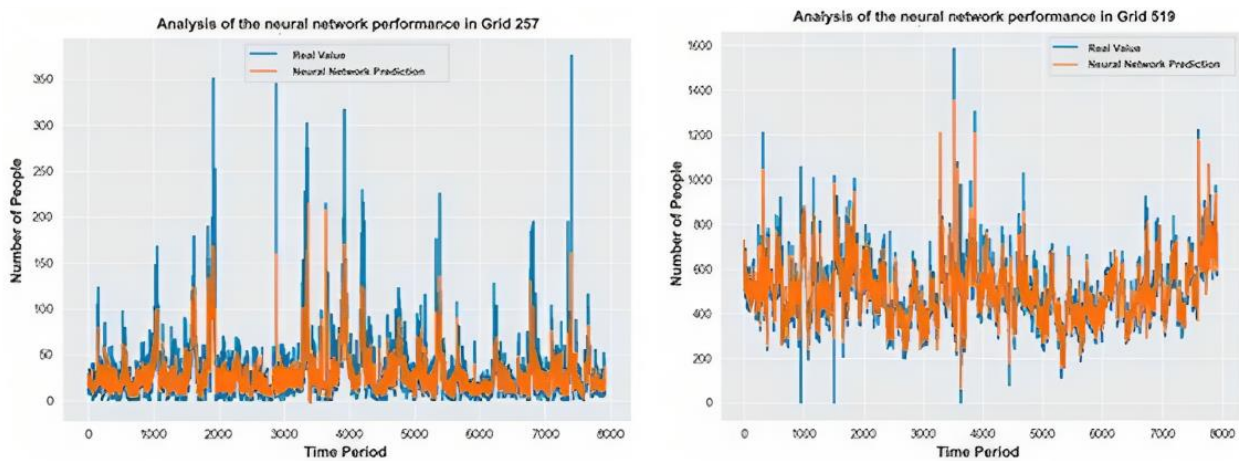


Figure 3. Performance of the network Grid 257 on left image and Grid 519 on right image.

Figure 4 serves as a representation of the outcomes observed when altering the dependent variable from the total quantity of terminals to the difference in the number of terminals between two consecutive periods. This adjustment in forecasting methodology was aimed at evaluating the networks’ capability to accurately identify spikes without being solely reliant on the absolute values of past periods. Encouragingly, the results were affirmative, indicating that the neural network effectively discerns spikes in the data even when considering relative changes between consecutive periods.

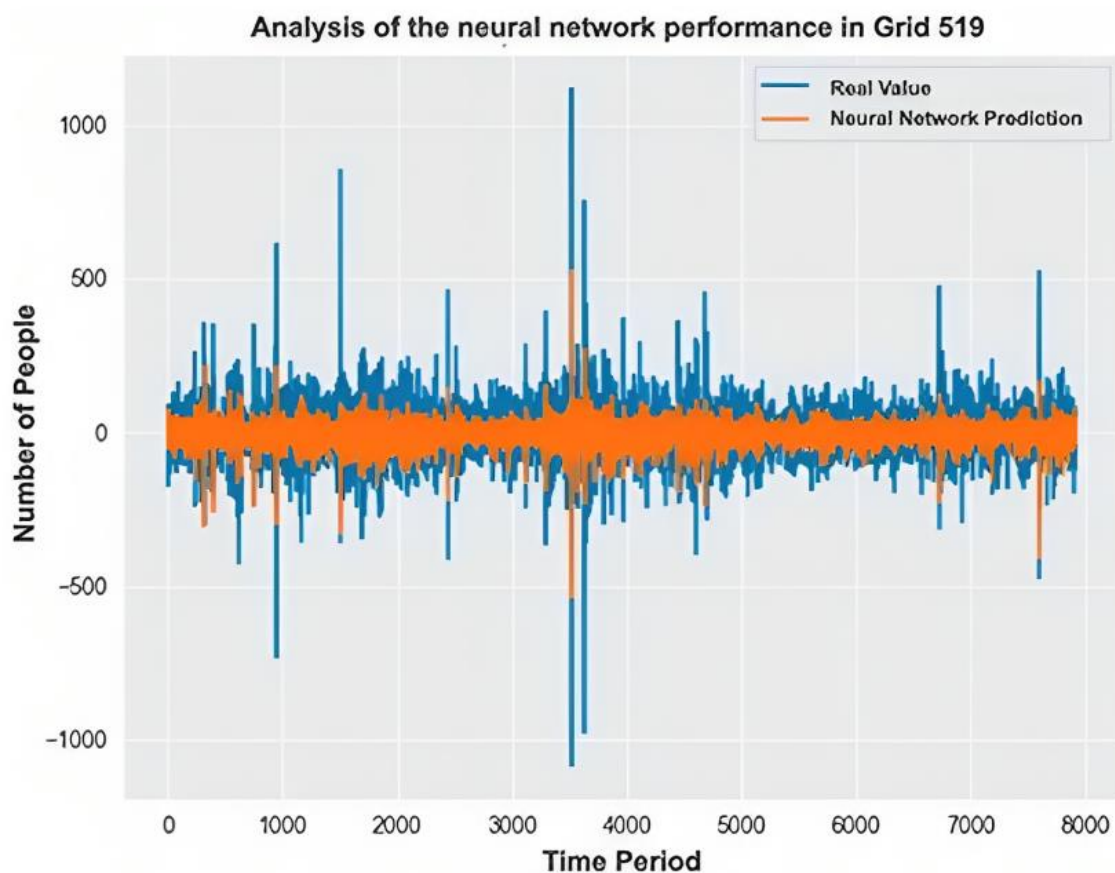


Figure 4. Performance of the network in grid 519 (change in dependent variable).

As previously highlighted, LSTM solely incorporates the temporal aspect of the dataset; however, it has been established that neighboring values also exert influence. In an effort to integrate the spatial dimension into the forecasts and enhance their accuracy, we augmented each data point with the values from neighboring areas. Despite this attempt, the obtained results fell short of expectations, encountering a pronounced issue of overfitting, leading to a notable deterioration in forecast accuracy.

In future research endeavors, it is imperative to explore advanced deep learning methodologies that seamlessly integrate both spatial and temporal components within a singular neural network. Approaches such as adapting the LSTM network to incorporate spatial information through Graph Convolution [37] merit investigation. This innovative direction holds promise for overcoming the challenges faced in our attempt to harmonize temporal and spatial considerations, potentially yielding more robust and accurate forecasting models.

5. People's Concentration Dashboard

Recognizing the critical importance of real-time predictions for city management, the outcomes of this study have been translated into a dashboard. This dashboard provides a platform for local experts to monitor and assess people concentration over time, facilitating informed decision making and responsive city management.

The identification of Lisbon's nightlife areas, along with the method to make forecasts, allows for the creation of the dashboard presented in this section, intended to help the Lisbon city council in its decision-making process. Given the availability of LSTM networks developed in Python, we decided to use this programming language to construct the dashboard. The Dash framework (<https://dash.plotly.com> (accessed on 1 December 2023)) was used to aid the process.

The homepage (Figure 5) allows for the visualization of a heatmap showing the density of people in the nightlife areas, updating itself every 5 min. The movement patterns of people are displayed using a color spectrum ranging from green, representing areas with a low number of people, to yellow and red, representing moderate and high concentrations of individuals, respectively.

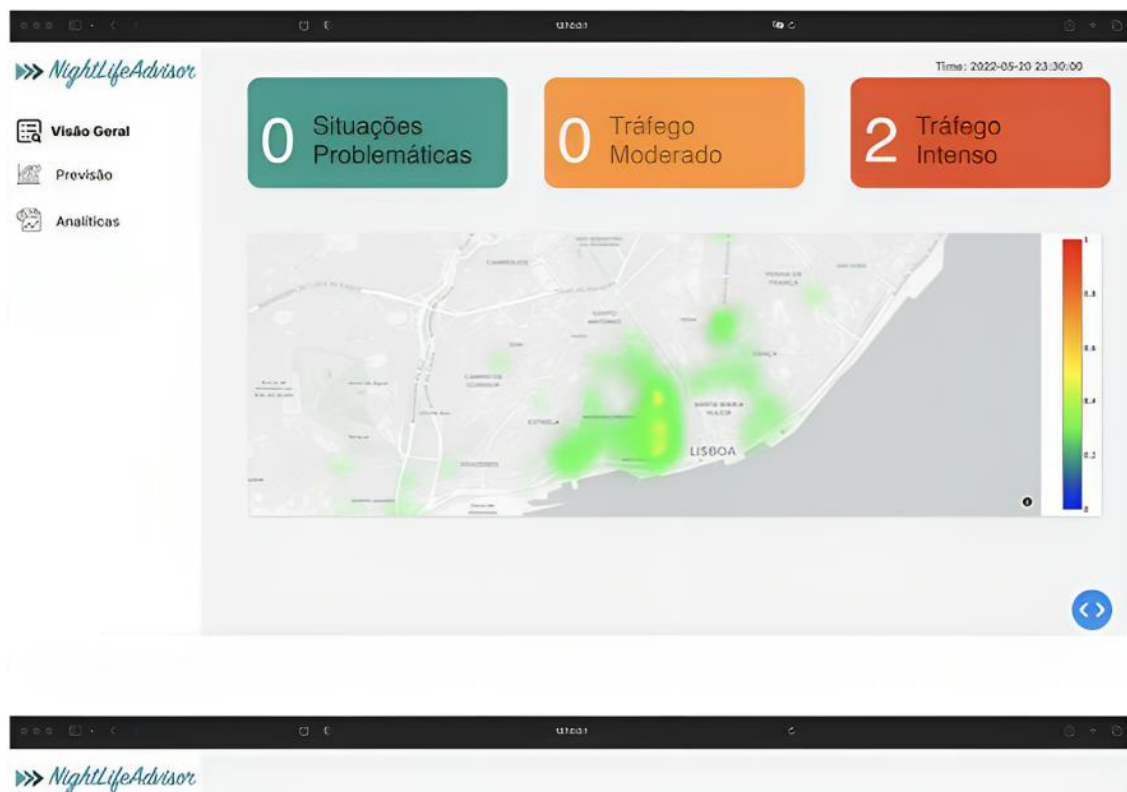


Figure 5. People’s concentration dashboard.

The dashboard illustrates people concentration in both temporal and spatial dimensions, offering city council experts the ability to navigate seamlessly through time and space for a comprehensive understanding of city demographics. Positioned above the heatmap are three distinct sections, arranged from left to right: the count of problematic situations (Situações Problemáticas) identified by the algorithm as abnormal; the instances characterized by moderate people traffic (Tráfego Moderado); and the occurrences featuring high people traffic (Tráfego Intenso).

Regarding the forecasts, the dashboard imports the LSTM networks developed with TensorFlow, loads them, and offers near-real-time predictions of the number of people in a given area for the next 5 min, with the potential to extend the forecasts to longer periods.

The dashboard contains a map of Lisbon’s nightlife areas, and the user may select the desired area to see the forecast of the number of people and past activity (see Figure 6). The dashboard also offers summaries and a detailed analysis of the movements in the last 12 or 24 h, as well as in the past 7 or 14 days.

Currently, the dashboard is installed in an ISCTE-IUL web server (see simulation at <https://t.ly/jIUqU> (accessed on 1 December 2023)). However, it has the potential to be integrated into the city council’s server and, with a live connection to the data, to offer near-real-time forecasts and analysis.

The enhanced dashboard offers intricate predictions, complete with dynamic visualizations showcasing the evolution of people’s concentration over both time and space (refer to examples in Figures 7 and 8). It furnishes users with pertinent insights crucial for

effective city management. Serving as a vital tool, it empowers authorities to anticipate future concentrations through leveraging historical event data.



Figure 6. People’s concentration prediction by zone and time.

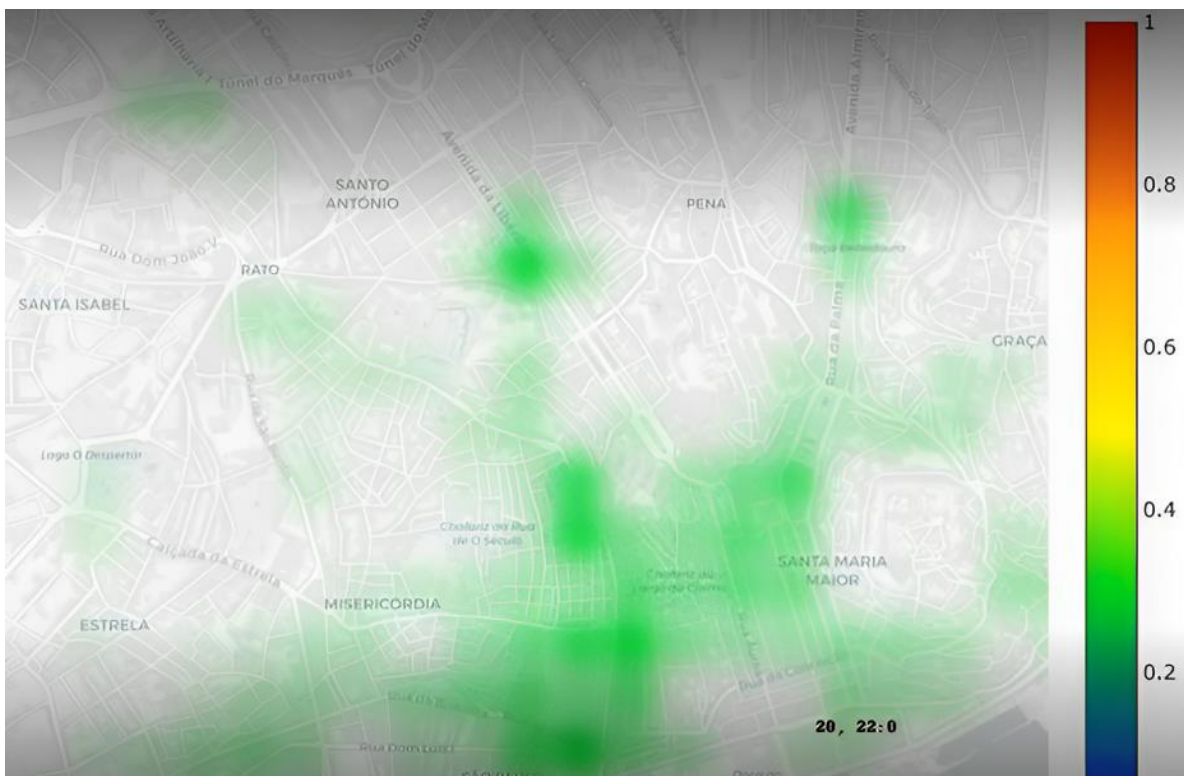


Figure 7. People’s concentration dashboard on 1 June at 7 p.m. in Lisbon.



Figure 8. People's concentration dashboard on 12 June at 8 p.m. in Lisbon.

These prediction animations can be valuable tools for urban planners, policymakers, and researchers to gain insights into various aspects of city life and management, through the following benefits:

- **Traffic Management:** Understanding the ebb and flow fluctuation patterns of people throughout the day can aid in optimizing traffic management and public transportation schedules. It can help identify peak hours, congestion-prone areas, and the need for infrastructure improvements.
- **Resource Allocation:** By tracking population concentration, city officials can allocate resources more efficiently. This includes emergency services, healthcare facilities, educational institutions, and more.
- **Urban Planning:** Planners can use these prediction animations to assess the demand for housing, commercial spaces, and recreational areas in different parts of the city. This can inform zoning decisions and urban development strategies.
- **Public Health:** Monitoring population density can be crucial during disease outbreaks or pandemics. It helps in identifying areas with high infection risks and implementing targeted measures like testing centers and vaccination clinics.
- **Safety and Security:** Law enforcement agencies can use such data to understand where and when crowds gather, allowing them to deploy personnel more effectively during events or emergencies.
- **Environmental Impact:** Population density data can be used to study the environmental impact of urbanization, such as energy consumption and pollution levels.
- **Economic Analysis:** Understanding when and where people are concentrated can inform businesses about peak hours for sales and service demands.

The implications of our research are significant. Urban planners, policymakers, and city officials can utilize the insights provided using our approach to make informed decisions regarding transportation infrastructure, emergency-response planning, crowd management, and urban design. By embracing these predictive capabilities, cities can strive for greater efficiency, sustainability, and livability.

Importantly, our work underscores the importance of respecting individual privacy and adhering to ethical considerations. We emphasize the use of anonymized and aggregated data to ensure that the benefits of our predictive models are achieved without compromising user privacy rights or violating regulatory norms.

While our paper contributes a novel predictive framework using neural networks, it also opens avenues for future research. Exploring the integration of additional data sources, such as weather patterns, social media activity, and public events, could further enhance the accuracy of predictions. Additionally, refining the models to accommodate evolving urban dynamics and exploring the potential of real-time predictions could yield even more actionable insights for smart city initiatives.

Our predictive framework employs machine learning techniques, including clustering and spatiotemporal analysis, to uncover patterns in human movement and concentration across various city regions. Through rigorous evaluation using real-world mobile phone data, we have shown the accuracy and effectiveness of our models in anticipating areas of high concentration and predicting people's movements during different times and circumstances.

Expert feedback indicates consistently high precision, particularly with a one-hour time window.

6. Conclusions

In this paper, we have presented a neural network approach to predicting people's concentration and movements within a smart city environment using mobile phone data provided by telecommunication operators. This approach improves results through addressing the critical challenge of understanding the dynamic urban dynamics that influence resource allocation, urban planning, and overall quality of life. By leveraging the ubiquity of mobile phones and harnessing the power of aggregated and anonymized data, we have demonstrated the potential of data-driven insights to revolutionize the field of smart cities.

In conclusion, this paper represents a substantial step forward in the realm of smart cities by demonstrating the viability of predicting people's concentration and movements using mobile phone data from telecommunication operators. As cities continue to grow and evolve, the data-driven insights presented here have the potential to shape urban planning strategies, optimize resource allocation, and ultimately create smarter and more sustainable urban environments for current and future generations.

Author Contributions: Conceptualization, B.F. and L.E.; methodology, L.E.; software, M.N.; validation, J.A.A., data curation, L.E. and M.N.; writing—original draft preparation, B.F. and J.C.F.; writing—review and editing, J.A.A. and J.C.F.; supervision, J.C.F. and B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundação para a Ciência e Tecnologia under Grant [UIDB/00315/2020]; and by the project “BLOCKCHAIN.PT (RE-C05-i01.01—Agendas/ Alianças Mobilizadoras para a Reindustrialização, Plano de Recuperação e Resiliência de Portugal” in its component 5—Capitalization and Business Innovation and with the Regulation of the Incentive System “Agendas for Business Innovation”, approved by Ordinance No. 43-A/2022 of 19 January 2022).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are available based on request and a NDA signature from Lisbon Municipality. Simulation available at <https://t.ly/jIUqU> (accessed on 1 December 2023).

Acknowledgments: EEA Grants Blue Growth Programme (Call #5), Project PT-INNOVATION-0069-Fish2Fork.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, X.; Gong, P. Urban growth models: Progress and perspective. *Sci. Bull.* **2016**, *61*, 1637–1650. [CrossRef]
2. Hosseinali, F.; Alesheikh, A.A.; Nourian, F. Agent-based modeling of urban land-use development, case study: Simulating future scenarios of Qazvin city. *Cities* **2013**, *31*, 105–113. [CrossRef]
3. Tayyebi, A.; Pijanowski, B.C.; Tayyebi, A.H. An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran. *Landsc. Urban Plan.* **2011**, *100*, 35–44. [CrossRef]
4. Wirth, R.; Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 11–13 April 2000; Volume 1, pp. 29–39.
5. Elvas, L.B.; Marreiros, C.F.; Dinis, J.M.; Pereira, M.C.; Martins, A.L.; Ferreira, J.C. Data-Driven Approach for Incident Management in a Smart City. *Appl. Sci.* **2020**, *10*, 8281. [CrossRef]
6. Elvas, L.B.; Gonçalves, S.P.; Ferreira, J.C.; Madureira, A. Data Fusion and Visualization towards City Disaster Management: Lisbon Case Study. *EAI Endorsed Trans. Smart Cities* **2022**, *6*, e3. [CrossRef]
7. Zhang, D.; Ge, Y.; Wu, X.; Liu, H.; Zhang, W.; Lai, S. Data-Driven Models Informed by Spatiotemporal Mobility Patterns for Understanding Infectious Disease Dynamics. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 266. [CrossRef]
8. Ma, Z.; Zhang, P. Individual mobility prediction review: Data, problem, method and application. *Multimodal Transp.* **2022**, *1*, 100002. [CrossRef]
9. Li, X.; Pan, G.; Qi, G.; Li, S. Predicting Urban Human Mobility Using Large-Scale Taxi Traces. 2011. Available online: <https://link.springer.com/article/10.1007/s11704-011-1192-6> (accessed on 1 December 2023).
10. Zhou, Z.; Yang, K.; Liang, Y.; Wang, B.; Chen, H.; Wang, Y. Predicting collective human mobility via countering spatiotemporal heterogeneity. *IEEE Trans. Mob. Comput.* **2023**, *22*, 4044–4055. [CrossRef]
11. Yao, Y.; Zhang, H.; Defan, F.; Chen, J.; Li, W.; Shibasaki, R.; Song, X. Modifiable Areal Unit Problem on Gridded Mobile Crowd Sensing: Analysis and Restoration. *IEEE Trans. Mob. Comput.* **2022**, *22*, 4044–4055. [CrossRef]
12. Phithakkitnukoon, S.; Horanont, T.; Di Lorenzo, G.; Shibasaki, R.; Ratti, C. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In Proceedings of the Human Behavior Understanding: First International Workshop, HBU 2010, Istanbul, Turkey, 22 August 2010; Proceedings 1. Springer: Berlin/Heidelberg, Germany, 2010; pp. 14–25.
13. Toole, J.L.; Herrera-Yaque, C.; Schneider, C.M.; González, M.C. Coupling human mobility and social ties. *J. R. Soc. Interface* **2015**, *12*, 20141128. [CrossRef]
14. Wang, J.; Kong, X.; Xia, F.; Sun, L. Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD Explor. Newsl.* **2019**, *21*, 1–19. [CrossRef]
15. Kothari, P.; Kreiss, S.; Alahi, A. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 7386–7400. [CrossRef]
16. Giffinger, R.; Fertner, C.; Kramar, H.; Meijers, E. City-ranking of European medium-sized cities. *Cent. Reg. Sci. Vienna UT* **2007**, *9*, 1–12.
17. De Montjoye, Y.-A.; Hidalgo, C.A.; Verleysen, M.; Blondel, V.D. Unique in the crowd: The privacy bounds of human mobility. *Sci. Rep.* **2013**, *3*, 1376. [CrossRef] [PubMed]
18. Hong, S.K.; Kim, K.Y.; Kim, T.Y.; Kim, J.H.; Park, S.W.; Kim, J.H.; Cho, B.J. Electromagnetic interference shielding effectiveness of monolayer graphene. *Nanotechnology* **2012**, *23*, 455704. [CrossRef] [PubMed]
19. Calabrese, V.; Cornelius, C.; Cuzzocrea, S.; Iavicoli, I.; Rizzarelli, E.; Calabrese, E.J. Hormesis, cellular stress response and vitagenes as critical determinants in aging and longevity. *Mol. Aspects Med.* **2011**, *32*, 279–304. [CrossRef] [PubMed]
20. Nikitas, A.; Thomopoulos, N.; Milakis, D. The environmental and resource dimensions of automated transport: A nexus for enabling vehicle automation to support sustainable urban mobility. *Annu. Rev. Environ. Resour.* **2021**, *46*, 167–192. [CrossRef]
21. Zhang, X. Sustainable urbanization: A bi-dimensional matrix model. *J. Clean. Prod.* **2016**, *134*, 425–433. [CrossRef]
22. Gounaridis, D.; Chorianopoulos, I.; Koukoulas, S. Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: The case of Athens. *Appl. Geogr.* **2018**, *90*, 134–144. [CrossRef]
23. Yang, J.; Guo, A.; Li, Y.; Zhang, Y.; Li, X. Simulation of landscape spatial layout evolution in rural-urban fringe areas: A case study of Ganjingzi District. *GIScience Remote Sens.* **2019**, *56*, 388–405. [CrossRef]
24. Kantakumar, L.N.; Kumar, S.; Schneider, K. SUSM: A scenario-based urban growth simulation model using remote sensing data. *Eur. J. Remote Sens.* **2019**, *52*, 26–41. [CrossRef]
25. Tsagkis, P.; Bakogiannis, E.; Nikitas, A. Analysing urban growth using machine learning and open data: An artificial neural network modelled case study of five Greek cities. *Sustain. Cities Soc.* **2023**, *89*, 104337. [CrossRef]
26. Alghais, N.; Pullar, D. Modelling future impacts of urban development in Kuwait with the use of ABM and GIS. *Trans. GIS* **2018**, *22*, 20–42. [CrossRef]
27. Li, F.; Li, Z.; Chen, H.; Chen, Z.; Li, M. An agent-based learning-embedded model (ABM-learning) for urban land use planning: A case study of residential land growth simulation in Shenzhen, China. *Land Use Policy* **2020**, *95*, 104620. [CrossRef]
28. Ramachandra, T.; Sellers, J.M.; Bharath, H.; Vinay, S. Modeling urban dynamics along two major industrial corridors in India. *Spat. Inf. Res.* **2019**, *27*, 37–48. [CrossRef]
29. Chaturvedi, V.; de Vries, W.T. Machine learning algorithms for urban land use planning: A review. *Urban Sci.* **2021**, *5*, 68. [CrossRef]

30. Tsagkis, P.; Photis, Y.N. Using Gama platform and Urban Atlas Data to predict urban growth. The case of Athens. In Proceedings of the 13th International Conference of the Hellenic Geographical Society, Athens, Greece; 2018.
31. Keras. Keras 3: A New Multi-Backend Keras. Available online: <https://github.com/keras-team/keras> (accessed on 12 December 2023).
32. Rouault, E.; Warmerdam, F.; Schwehr, K.; Kiselev, A.; Butler, H.; Łoskot, M.; Szekeres, T.; Tourigny, E.; Landa, M.; Miara, I.; et al. "GDAL." Zenodo, 30 November 2023. Available online: <https://zenodo.org/records/10410302> (accessed on 22 December 2023).
33. Frias-Martinez, V.; Soguero, C.; Frias-Martinez, E. Estimation of urban commuting patterns using cellphone network data. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 9–16.
34. Givens, C.R.; Shortt, R.M. A class of Wasserstein metrics for probability distributions. *Mich. Math. J.* **1984**, *31*, 231–240. [CrossRef]
35. Felix, A.G.; Jürge, S.; Fred, C. Learning to Forget: Continual Prediction with LSTM. *Neural Comput* **2000**, *12*, 2451–2471. [CrossRef]
36. Staudemeyer, R.C.; Morris, E.R. Understanding LSTM—A tutorial into Long Short-Term Memory Recurrent Neural Networks. *arXiv* **2019**, arXiv:1909.09586. [CrossRef]
37. Forecasting Using Spatio-Temporal Data with Combined Graph Convolution + LSTM Model—StellarGraph 1.2.1 Documentation. Available online: <https://stellargraph.readthedocs.io/en/stable/demos/time-series/gcn-lstm-time-series.html> (accessed on 18 December 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.