Journal of Computer Science

Review

Integrating Data Mining Methods Across all Domains of a Smart City

Rushit Dave, Nuo Xu, Simon Arneberg, Naeem Seliya and Mounika Vanamala

Department of Computer Science, University of Wisconsin Eau Claire, USA

Article history
Received: 01-04-2022
Revised: 07-05-2022
Accepted: 12-05-2022

Corresponding Author:
Rushit Dave
Department of Computer Science, University of Wisconsin Eau Claire, United States
Email: rushitdave05@gmail.com

Abstract: As modern cities continue to develop, smart devices are being used to improve citizens' lives. As these devices become more sophisticated, the amount of information that they collect increases. A Smart City is a city that uses a large amount of generated data to constantly improve the services offered to the people that live there. Data mining techniques can be used to sift through the data and mine out meaningful patterns. Our research project focused on seven different disciplines within a Smart City, surveying the current state of research in each category. Smart Transportation is focused on decreasing congestion, increasing efficiency in public transportation, and improving the safety of pedestrians. Smart Healthcare is focused on modern healthcare monitoring systems and ambulance dispatch services. Smart Energy is focused on decreasing energy consumption and promoting green energy through WiFi thermostat optimization and a smart electric grid. Smart City Utilities is focused on creating algorithms to improve waste collection techniques and air quality. Smart City Planning is focused on land use and the placement of green spaces. Smart Networks and Privacy is focused on secure networks. Lastly, the Smart IoT (Internet of Things) Application is focused on next-generation networks like 5G.

Keywords: Data Mining, Smart City, Machine Learning, Big Data, IoT, Smart Sensor, Traffic Management System, Cloud Computing, Edge Computing

Introduction

Throughout all of history, humans have chosen to live in communities that provide them with safety, public services, and the sharing of resources. However, cities are becoming increasingly populated across the world, which offers both unique challenges and opportunities for growth. Traditional cities are structured with infrastructures separated into many separate domains. Each of these domains is independent of every other domain and each domain works to benefit people in that specific area of life. Services and systems like transportation, emergency response, and energy networks need to find their solutions to the unique problems that they each face. However, modern technological developments have changed the capacity of cities to offer services and systems that traverse traditional domain borders to provide better care for the residents within (Sodhro *et al*., 2019). In this study, we will examine the current state of research promoting the integration of a smart city using data mining techniques. In our research, we found that seven critical domains are being researched to create a truly smart city: Namely Smart Transportation, Smart Healthcare, Smart Energy, Smart City Utilities, Smart City Planning, Smart Networks, and Privacy, and Smart IoT Application.

More data is being collected today than at any other point in history and we are offered an extraordinary opportunity to harness this data to benefit people's lives. Using the current research on smart cities as a starting block, we can continue to develop solutions that draw on the interconnected nature of the modern city. This study outlines what we identified as the seven main subcategories of research and gives a summary of some of the recent important research in that field. A city cannot simply adopt Smart Energy strategies and call itself a smart city. The smart city of the future must find ways to integrate strategies from every single domain, sharing information across every sector of the city. Only then will we be able to achieve a truly smart city.

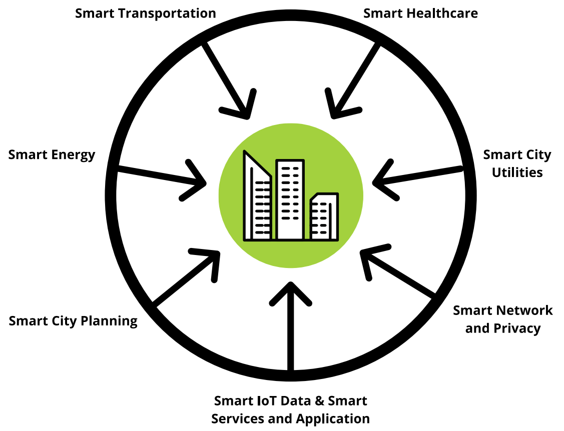

Fig. 1: Seven critical domains of a Smart City

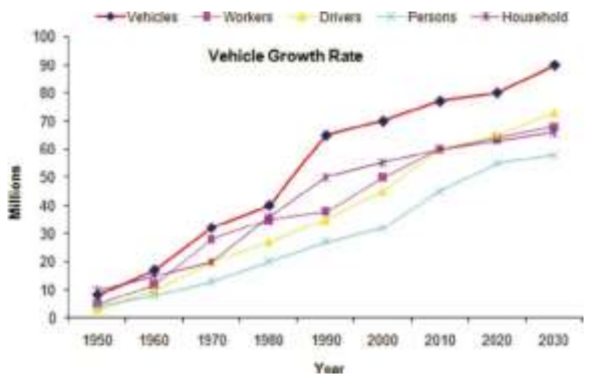

Fig. 2: Past and projected vehicle growth rate (Ragavan *et al*., 2021)

In this study, we will examine the Background and Literature, then present a Survey of Current Research breaking down existing research in each of the seven domains, then discuss our results in table format in Discussion and Analysis and lastly wrap up the paper in the Conclusion, followed by the list of our References.

Background and Literature

After collecting and analyzing 49 research and journal papers relating to data mining in smart cities, we were able to compile a list of seven critical domains of a smart city (Fig. 1), as listed in the introduction. It may be helpful to initially clarify the terms “smart city” and “data mining”.

Smart cities are rapidly becoming a reality in our ever-changing world. Paper (Wu *et al*., 2018) gives examples of actions that various cities around the world have already taken to become more aligned with the vision of a smart city, which is to use the wealth of data from interconnected sensors and devices to improve the wellbeing of citizens. These examples include the city of Dubuque, Iowa tracking each home's electricity and water consumption, the city of Los Angeles, California syncing up a grid of thousands of traffic lights to run more efficiently, and the city of Copenhagen building a new transportation network within one kilometer of all residential areas through analyzing housing data. Another paper (Pal *et al*., 2018) examines similar case studies around the world, as well as argues for a framework for sensing, storing, processing, and applying big data.

Data mining is a term that refers to various methods of “mining” useful information out of enormous amounts of data. Oftentimes, there is so much data that it would be impossible for a human to make any meaningful connections or inferences, which is where techniques such as machine learning algorithms and big data analytics come into play. According to (Al Nuaimi *et al*., 2015), we should be aware of the five V’s or big data management (volume, velocity, variety, variability, and value) whenever assessing a new data mining technique. This study also calls for more structure and standardization in data collection, which would allow for information to be mined more consistently across the world.

Survey of Current Research

This is the portion of the paper in which we will summarize important papers in each of the seven domains. Each section will incorporate relevant papers we reviewed to create a cohesive argument for their implementation in a smart city. We will begin this portion by discussing Smart Transportation.

Smart Transportation

Transportation is essential to the implementation of a smart city. As worldwide populations continue to increase, more and more people are migrating to cities, which means cities are becoming more and more crowded (Fig. 2). Traffic congestion is not merely inconvenient for the individual. Rather, congestion is harmful to society because of excess fuel consumption, increase in accidents, increase in poor driving behaviors such as road rage, health problems caused by stress, and environmental waste. In addition to this, there is an enormous economic cost when workers lose out on thousands of hours of work. It is for these reasons that we believe efficient transportation to be a critical element of a smart city.

Finogeev *et al.* (2020) Professional data collection and processing system infrastructure (measurement equipment, photoelectric radar complex, road surveillance cameras, etc.) can be used to track the elements of road transportation. This program is an automatic receiving process to obtain sensory data by polling the photoelectric radar system, video surveillance cameras, and other sensor nodes located at or near the highway. External data collection and processing system (road weather station, satellite navigation system, intelligent transportation system, etc.). An example is data vehicles and response systems from navigation equipment (navigators, transponders, GPS/GLONASS modules, ERA-GLONASS, eCall, EVAC, E911 systems, etc.). The program allows the identification of factors that increase the likelihood of a traffic accident depending on the current road environment, weather and road conditions, road surface, time of day, road traffic, etc. The collection and processing of text, geographic space, photos, and video data from road-mobile communication users and third-party observers.

Photographic radar complexes, cameras, and related networks have detected data transmission nodes in the region where they are located. The system consists of geospatial tags and cartographic names of objects. The information comes from the complex and the central repository presents metadata containing its attributes. The latter includes the identification number of the complex, the longitude, and latitude of the object's location, name and complex description (brand, model, etc.), the address or orbit of the postal object and the photo of the object with text description and hash information (digital signature) Used to identify the data from the complex, the location of the area code object, the speed limit of the control section, the number of lanes and the IP address.

An analysis of the abnormal data displayed on the chart shows. After April 17, the number of people dropped sharply and road accidents occurred at all complex facilities at the same time. Meteorological data is collected to determine the cause of the abnormal decrease. It was recorded by Cordon-Temp equipment on the M-5 (Ural) route after April 15^{th}. These days, unfavorable weather conditions (rain, snow, low temperature) cause icing. Next, within a few days, a push message was sent to the mobile communication with the predicted result warning the driver to be careful. As a result, the road traffic volume was reduced by 20% and the average speed of the controlled section of the road, in which the camera radar was installed in the complex, was reduced by 40%.

In the process of spatial analysis, the identification of similar parts of the road traffic infrastructure is based on the number and types of road traffic accidents. Clustering allows you to identify the most critical and emergency sections of the road and display them in the area of the driver's cartographic basis distinguished by the color of dangerous goods. The information is also sent to the mobile phone and uploaded to the user's communication facility application. The mobile application agent warns drivers to be more cautious about the need to enter critical areas. Traffic risk accidents can be reflected by frequency and severity. Implement an active monitoring system to reduce the risk of accidents and disasters and to ensure the safety of human life in the urban environment.

There are four main methods that a smart city could use to monitor traffic, (Putra and Warnars, 2018). These four methods are motion sensor, ultrasonic sensor, speed sensor, and passive infrared sensor monitoring. A smart city needs to know all the tools at its disposal because the implementation of an intelligent traffic monitoring system could help to minimize traffic accidents that result from human error.

To make transportation networks more efficient, another paper (Ragavan *et al*., 2021) proposes an intelligent traffic control system that uses an Internet of Vehicles (IoV) to communicate between vehicles. A map of India is split into many segments of equal size and an Ant Colony Algorithm (ACO) is used to determine the optimal path for vehicles to travel on. The ACO algorithm is based on the activity of ants, who put down a pheromone of varying intensity along a pathway. Gradually, more and more ants follow the optimal routes, creating an efficient travel network. This is the basic idea for the ACO algorithm, except the traveling paths of vehicles are simulated and can be used with IoV data. The IoV is reliant on two types of communication: Vehicle-to-infrastructure communication and vehicle-to-vehicle communication.

Traffic density estimation is done using a Support Vector Machine (SVM). The SVM examines video snapshots of roads, considering information about the location and time of the images. After this, it classifies the images by how many cars there are present in the video and then calculates the density of cars for that particular stretch of road. Overall, the path selection using the ACO algorithm performed better at finding the optimal route than existing algorithms in the simulations.

In addition to improving traffic flow for citizens, improved road efficiency and mapping could help immensely with emergency response time. This (Liu and Wang, 2019) paper outlines a few methods to do this. Congestion is especially harmful in the case of emergency vehicles, which happens to be the most time-critical endeavor of traffic flow. The goal is to create a method to discover safe and fast routes for emergency response vehicles to minimize congestion in cities. Utilizing the Internet of Things with data mining techniques would be the most beneficial method of helping cities in emergencies. In the case of evacuations, personnel should be given RFID (radio frequency identification) bracelets to quickly summarize the flow information of people and vehicles. This would give officials valuable information about where people are headed, where people are evacuating, and where dangerous congestion might be occurring. In addition, the study also analyzes the application of some of these technologies in the city of Qingdao.

Because some of the proposed data mining techniques have not been implemented in the real world with historical accuracy, the authors advise that we use them not as firm directives, but rather as informative input to officials who are making the decisions. The implementation of these methods will greatly improve rescue efficiency and quality, according to the author's calculations. Future research includes more accurate data collection, as well as applying these techniques to other types of disasters in smart cities.

As we argued earlier, there is an enormous need for adaptive traffic management in a smart city. For example, (Yousef *et al*., 2019) propose novel history-based traffic management algorithms to predict the congestion levels of streets in a city based on data collected year-round. Using this information, the red/green times for each direction can be calculated. The authors test their proposed algorithm using MATLAB and a traffic simulator SUMO. In addition to one intersection, they also tested the model on a grid of 4 × 4 separate intersections and tested it compared to existing algorithms. They found that the proposed algorithm optimizes traffic flow up to 18% more than standard traffic management.

All-year traffic history allows the algorithm to predict the levels of traffic based on previous years before any current data is even collected. The proposed algorithm performed better in busy traffic conditions, but slightly worse in normal traffic conditions. Thus, the use of history is helpful in a certain situation and less helpful in others, so it will take further analysis to determine when and how much to use each method. The proposed method can be generalized to real-life intersections. For future research, incorporating road conditions and different intersection models into the model would help achieve the most useful results.

In addition to car transportation, some research is being done to improve the efficiency and capabilities of public transportation systems. Sari Aslam *et al*. (2021) This study outlines a method to predict a traveler's trip purpose, given smart card data and points-of-interest data. This neural was tested in the real world (in London) and has proven to be effective in predicting the purpose of a trip. But smart cities do not only need to deal with cars and buses. They also need methods to deal with humans. A paper proposed a new model for detecting trajectory outliers in human behavioral data (Belhadi *et al*., 2021). This is useful in traffic situations because it is used to model pedestrian behavior within a smart city. Using deep learning algorithms, the model was able to take mapped data of human behaviors, a process that data and identify collective abnormal human behavior.

Smart Healthcare

We turn our attention now to the well-being of citizens residing in a smart city. Data mining techniques can be used to improve the health of the general population, both through reactive and proactive measures. Nowadays, more and more smart devices are designed and manufactured and with continuous upgrading and technological innovation, these smart wearable devices have entered the daily life of the public. In a sense, these smart wearable devices that integrate various functions are mobile sensors. They can not only monitor human gestures such as recognizing movement patterns but also alert people to relevant medical emergencies at critical moments. According to (Rayan *et al*., 2019), Smart Healthcare could provide people with benefits such as patient monitoring, aid in self-diagnosis, and even early detection of diseases. To achieve these fantastic ends, however, there are a few obstacles that we must first overcome. Rahman *et al*. (2021) This study outlines the importance of smart healthcare data security, data privacy, and social acceptance of deep learning solutions.

The reason why smart medical systems are smart is that the intervention of a large number of smart devices, namely biosensors, generates a large amount of data for risk monitoring and disease assessment. These sensor data are usually integrated into one or more public health service agencies in the user's area. As a specific example, (Han *et al*., 2020) propose a long-term posture-recognition method using sensors and LoRa technology. The method uses Random Forest to select the key features that the LoRa nodes require. The proposed posture recognition system consists of a posture information sensor module, wireless transmission module, human posture recognition module, and user interface module. In modern urban life, more and more people are beginning to pay attention to the health of themselves and their families and the ability to detect posture can lead to healthier lifestyles for the people living in them. Even in some extreme cases, posture detection sensors can quickly detect falls. Based on (Pfortmueller *et al*., 2014) serious injuries such as fractures and traumatic brain injury occur in approximately 10% of falls. Therefore, we believe that smart biosensors will be widely used in different scenarios in the future. Especially in the context of the COVID-19 pandemic, the threat of infectious diseases to public health security has put medical institutions under unprecedented pressure. We hope to explore the expansion of the application of medical care equipment in the future smart city to establish a complete and scalable intelligent medical system framework. The framework also needs to address the telemedicine needs of patients with communicable diseases when they are self-isolated.

In (Enler *et al*., 2020) we discover an extensible, maintainable Service-Oriented Architecture (SOA) (Fig. 3). However, in a large network of smart devices, the data often presents different kinds. Thus, researchers introduced a public data API to classify and process different kinds of data. In addition to the smart wearable devices mentioned above, smart furniture will also popularize health monitoring functions in the future, such as smart scales, smart treadmills, etc. These smart devices generate different types of health data. Therefore, the API introduced by this architecture (SOA) can effectively classify different data for subsequent higher data analysis efficiency. This truly realizes the scalability of the smart medical system.

Disease monitoring and prevention are the most effective ways to improve health. However, this does not prevent the occurrence of disease. Therefore, remote consultation and emergency medical response are also important components of the intelligent medical system. In (Gera *et al*., 2021), smart biosensors collect key biological data such as pulse rate, blood pressure, blood sugar, heart rate, etc. Sensors send data to the cloud, where critical threshold alerts are set and machine learning models are used for initial screening and data analysis. Cloud storage is also connected to specific medical institutions so that doctors can remotely diagnose and open prescriptions and all case and prescription information are also uploaded to the cloud for pharmacies. The application of biosensors and cloud data has made full-process telemedicine a reality (Fig. 4). General visits, prescriptions, drug orders, and appointments for pathological tests can all be done at home. Smart healthcare can not only improve the medical experience but also reduce the waste of medical resources and better monitor disease progression and treatment effects. For patients with communicable diseases, telemedicine reduces the risk of infecting others when visiting a doctor.

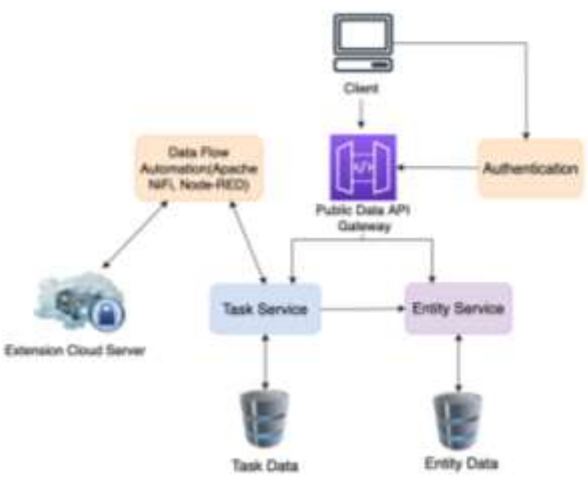

Fig. 3: Service-oriented architecture of smart healthcare

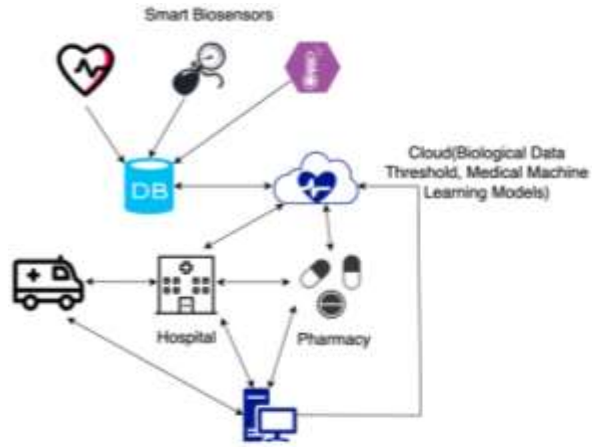

Fig. 4: Smart telemedicine and emergency medical services

After solving the patient's daily remote consultation needs, we turn our perspective to emergency medical services. Patients who need emergency medical services can be said to have the highest priority and the most difficult treatment in the entire medical process. They often suffer from sudden illnesses such as sudden death or accidents such as car accidents. This requires the emergency response time of ambulances to be shortened as much as possible. Existing traditional emergency medical services require patients to provide accurate addresses for medical events. However, in many cases, the parties or their families cannot provide accurate addresses due to illness or emotional breakdown. This resulted in the inability of emergency personnel to arrive quickly to treat the patient. Therefore, the intelligent ambulance management system proposed by (Akca *et al*., 2020) shortens the time in each process and aims to arrive at the scene as quickly as possible and send the patient to the emergency center. This management system has three different ports, namely dispatch center, ambulance, and patient. Registered users of the system are required to fill in personal information such as personal address and contact information in advance. When an ambulance is called, the system automatically retrieves location information from the background and sends it to the dispatcher and ambulance. The dispatch center interface displays the location and other personal information of the patient requesting an ambulance and the nearest ambulance and hospital. The ambulance interface can link with the dispatch center to display the patient's location and the nearest medical institution. In (Akca *et al*., 2020), the researchers tried to filter the nearest hospital location to the patient from the database and choose the fastest transportation method, using the Google API to make a matrix distance request. Determines the latitude and longitude of the largest circular distance between given spheres by the half-sine formula. In addition, the system can adjust the traffic lights on the ambulance's path, saving time through the intersection by moving the GPS signal to turn the nearest light on the ambulance's path to green until the ambulance passes.

Smart Energy

More infrastructure always comes at a cost: Energy. As a city expands, its energy usage continues to grow with it. Some of this energy is consumed by very necessary systems, but much of a city's energy usage is an extraneous waste. This waste could be prevented if proper measures are put in place. It is the job of data mining techniques to sift through the mountain of data to figure out exactly how to optimize the energy network of a smart city. According to (Babar *et al*., 2021), there are three main areas of focus for smart energy: Energy management, data processing, and service management. By analyzing and automating, a green energy grid is within grasp.

From the existing research field, we find that there are a large number of machines in the manufacturing factories that are idle. On average, computers are idle 16% of the time during an eight-hour work cycle, according to reports from aviation industry suppliers. Only shutting down the machine for at least 13% of its operating time during these inactive hours avoids wasting energy. Pei *et al*. (2020) It is well known that providers of modern Internet services usually need to establish multiple data sites in different regions to provide services such as Web applications and these data centers require a lot of power to power the basic operation of the data nodes and the cooling system. Currently, the world's electricity supply mainly comes from fossil fuel plants such as coal-fired and gas-fired power stations. About two-thirds of the world's electricity is generated by burning fossil fuels. As more and more countries are aware of environmental issues such as global warming, investment in renewable energy generators is also increasing. More and more network providers are opting for renewable energy generators to power data centers because of lower construction and electricity costs. However, compared with brown power such as coal-fired power and gas-fired power, renewable green energy such as wind, hydro and solar energy is unreliable, which brings huge challenges to the stability of data center operation.

We find a study introducing an intelligent and Sustainable Probability Distribution that can be used in a Hybrid Genetic approach (SSPD-HG) to reduce energy consumption and minimize the total completion time of a single machine in a smart city machine interface platform. Among them, to reduce the energy-delay of the data center, we need to carry out continuous energy management of the data center to make the power resource allocation of the data center reasonable and efficient (Fig. 5). It is worth noting that when a Hybrid Genetic approach (SSPD-HG) is used to control the energy usage of data centers, delay tolerance and opportunistic scheduling based on spatial load balancing and workload control can improve the utilization efficiency of green energy and minimize its cost.

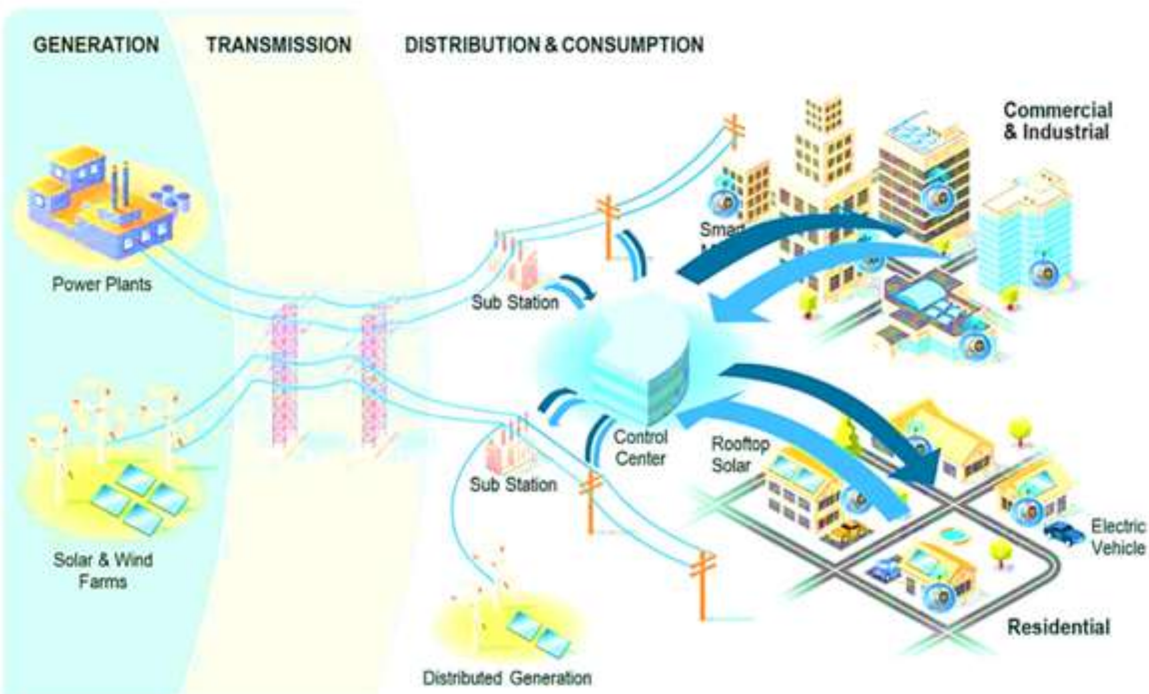

Fig. 5: Large-scale electric energy collected from the renewable energy sources (Pei *et al*., 2020)

The main principle of this method is to reduce the running time of a single machine in the data center. By monitoring and analyzing the real-time data bandwidth rate between users and the data center to adjust the machine working time and energy distribution, it can achieve energy-saving and load management at the same time.

Except for the new Internet enterprises, the energy consumption of urban basic industries occupies the main position of all energy consumption. Scientists discovered a steel industry-based data mining technique for analyzing and predicting data from IoT systems. Among them, they use existing statistical algorithms to predict industrial energy consumption, such as general linear regression, classification and regression trees, support vector machines with radial basis kernels, etc. After the prediction is completed, the root means square error, mean absolute error, and coefficient of variation should be used to measure the prediction efficiency of the model. Many scientists have examined energy consumption models using data mining techniques and it is worth noting that machine learning algorithms are very convenient and useful to verify the model after the model is established. VE *et al*. (2021) The researchers used smart meters to collect the energy consumption of machinery and equipment in the steel industry every fifteen minutes and to collect and store additional information about energy consumption in a cloud system. One year's energy consumption data of the steel industry is divided into a training dataset and a test validation dataset and a classification method are used for function and regression training to generate a predictive model. Scientists also use different performance measures to compare the efficiency of regression models such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Variation (CV). In this case, the results show that the SVM model has the highest stability and prediction accuracy.

In daily life, energy is the basic living need of every household. Prediction and management of energy consumption for each household can effectively reduce energy costs and waste caused by excess energy. The researchers found that upgrading low-income housing to middle-income household efficiency would reduce excess energy by 68%. Surveys conducted by technicians in the past have found that many factors affect the energy consumption of residential buildings, such as weather, building geometry, heating, ventilation, thermal insulation materials, etc. To study building energy consumption models, scientists conducted leak tests using infrared imaging to document wall insulation and assess the thermal efficiency of water heating systems. There are many more data-gathering investigations like this, thus it might be costly. The U.S. Department of Energy estimates the cost of a detailed energy audit to range from $0.12 to $0.503 per square foot. Therefore, smart sensors are of great significance for the realization of modern smart cities. It can improve the efficiency of data perception and greatly reduce the cost. In this study, researchers trained a machine learning model to effectively predict residential energy efficiency by installing a smart WIFI thermostat and measuring and collecting data in the form of a power spectrum against other data such as weather data. The energy data collected and transmitted through the thermostat can be used to develop a general predictive model of household energy characteristics. The researchers periodically classified outdoor temperature data into discrete temperature bands and used statistical methods to determine the probability density of outdoor temperature. The inverse energy model is obtained by further taking the average outdoor temperature as the input for the whole prediction period. The researchers applied and validated different machine learning algorithms on a computer including complete training and testing datasets. The results show that the house controlled by the Gradient Boosting Machine (GBM) based power spectral density prediction model has the highest energy thermal efficiency.

Building datasets and predictive models are not the ultimate goals of scientists. We hope that these models will serve as a tool for cities to use data analysis to predict demand and ultimately help shape more effective policies. From a commercial point of view, intelligent management of data center energy consumption can not only reduce the construction and operation costs of network providers but also reduce the use costs of consumers. To achieve the goal of intelligent management of modern urban energy networks, it is not enough to only focus on the energy consumption of industrial buildings and private houses. Energy storage, energy load transfer, and energy consumption analysis forecasts for the entire city are all aspects that we need to consider.

On-site energy storage is a difficult problem in traditional urban power grids and many scientific research teams try to overcome the problem that excess power is difficult to store. In the future, the smart grid in the Internet of Things environment is expected to completely solve the problem of on-site energy storage. Golpîra and Bahramara (2020) Researchers propose an IoT-based energy management framework that can be applied to integrated power distribution networks in smart cities. This framework allows for both centralized and decentralized decision-making mechanisms, which are also known as cloud computing and edge computing. And each decision-maker can solve and optimize their problems independently and the unique feature of this system is that decisions can be shared between the core cloud and edge cloud. A Smart Grid (SG) applying this framework can intelligently control the production, transmission, and distribution of power generation equipment by collecting information. The distribution network consists of multiple feeders, each of which in turn consists of different loads, energy storage, wind turbines, and microgrids. Edge computing can determine the energy balance of each microgrid in the mobile photovoltaic generation load, which is solved by the microgrid aggregator. Cloud computing uses the distribution system operator to meet the energy balance of the distribution grid with the output power of wind turbines, optimize energy storage scheduling and conduct electricity transactions with the market promptly. The edge cloud model consists of 13,453 single equations, 9,073 single variables, and 2,880 discrete variables. The model core cloud consists of 1,513 single equations, 1,153 single variables, and 144 discrete variables. After cloud computing optimization, the power grid has increased the energy storage capacity. The distribution system operator buys more energy from the market at lower market prices to quickly charge energy storage devices and power the grid when electricity prices are high. In the smart grid based on IoT technology, the microgrid is directly managed by the microgrid aggregator in each edge cloud and the distribution system operator is responsible for managing the resource allocation of the core cloud. This hierarchical trial edge computing framework reduces the total operating cost of the urban power grid by about 14.28%.

Smart City Utilities

Smart City Utilities encompasses various areas of study, including waste collection, air pollution, and smart emergency response services. While each of these subcategories is vital on its own, we identified their similarities and grouped them under the name Smart City Utilities. As cities expand, certain negative consequences are unavoidable, such as waste and pollution. Although they may be largely unavoidable, there are still certain measures that a smart city can take to mitigate these factors.

With the process of urbanization and economic development, the daily operation of cities produces a large amount of garbage, including domestic garbage and industrial garbage (Fig. 6). Reasonable collection and disposal of garbage can reduce pollution and the production cost of certain products. Therefore, how to collect garbage quickly and efficiently has become an urgent problem to be solved. (Akbarpour *et al.*, 2021) The number of waste removal vehicles can be optimized to minimize the cost of transportation. Oralhan *et al.* (2017) We discovered an intelligent waste management system that analyzes container storage for temperature, carbon dioxide, and liquid levels by installing integrated sensors inside the waste container. At the same time, the researchers also synchronized the most efficient garbage collection path to the cellular smart tablet inside the garbage truck based on the ant colony algorithm. To improve garbage collection efficiency, the researchers hope that a single garbage truck dispatch can collect as many containers as possible. Therefore, the garbage containers are clustered by K-means data mining technology, and the 200 garbage containers are divided into 15 clusters. The garbage cans in each cluster have a similar garbage loading level. When the cluster triggers the key threshold, the garbage truck will carry out collect. In real life, garbage trucks are often responsible for collecting garbage in several blocks through a specific route at a fixed time. This can lead to issues like untimely garbage collection or wasting time on empty garbage containers. In, researchers allocate garbage trucks based on garbage container clusters derived from data mining techniques and the collection paths of these garbage trucks are also optimized by an ant colony algorithm. The application of the intelligent waste management system significantly reduces the fuel cost of garbage trucks, truck wear and tear, and labor costs. The overall savings in waste collection expenses are about 30%. Another IoT-based solid waste management system uses similar technology. The difference is that (Bharadwaj *et al.*, 2016) specifically introduced that the management system uses LoRa technology to send sensor data to the gateway and upload it to the cloud using the MQTT protocol. The advantage of using LoRa is that it can achieve long-distance data transmission with low power consumption. Meanwhile, garbage truck collection routes are based on Google API and garbage bin GPS coordinates to analyze the best route. Due to the improvement of waste recycling technology, the government is also formulating regulations to require people to complete the garbage sorting work when throwing garbage. Then comes the need for the classification of smart trash cans. Lokuliyana *et al.* (2018) The proposed garbage management system makes a more detailed distinction between the types of garbage cans and uploads the types and locations of garbage cans to the user interface so that people can find the nearest correspondingly classified garbage cans.

Having successfully tackled the problems of garbage collection and garbage pollution, we will focus on another pollution that is not so easy to detect: Air pollution. Industrialization and the increase in the number of motor vehicles inevitably lead to a decline in air quality. Therefore, intelligent prediction methods for air quality can give people better advice on going out and help the government improve relevant laws and regulations. Because of the temporal and spatial diversity of weather, many air pollution sensors are installed (Zaree and Honarvar, 2018) to collect latitude and longitude, the main components of air pollution (ozone, particulate matter, carbon monoxide, nitrogen dioxide, sulfur dioxide) and meteorological data (temperature, humidity, air pressure, wind speed). The researchers used the K-Means algorithm to cluster the data output. After finding the optimal cluster, the researchers used RMSE and MSE statistical methods to evaluate the prediction model and finally came up with an air pollution prediction model. Building a predictive model is just to be able to detect air quality, the next step is to communicate air quality information to people. The Air Sense architecture mentioned in (Wray *et al*., 2018) can collect sensor data and deploy predictive models in the cloud and then send the data analysis results to a terminal such as a smartphone. Users can see air quality information in various areas of the city on the web interface and the architecture can also issue early warnings under extreme air quality. To improve the accuracy of predictive models, researchers (Honarvar *et al*., 2019) also tried neural networks and regression as the core of predictive models. Continuously improving prediction models can give people the most accurate travel advice in the future and help environmental regulators to prevent and monitor industrial pollution. Continuously improving prediction models can give people the most accurate travel advice in the future and help environmental regulators to prevent and monitor industrial pollution.

The main purpose of urban public utilities is to provide better services to citizens and maintain the basic operation of the city. In addition to the continuous urban pollution problem we mentioned, every city is experiencing various emergencies every day, such as crimes, disasters, accidents, etc. This requires cities to improve the emergency response mechanism and improve the level of emergency response services in the process of development. Emergencies are often uncertain and sudden. Therefore, the core of urban emergency response services is the rapid arrival of emergency response units on the scene. But in cities, the high number of motor vehicles can cause traffic jams, especially during rush hour. Traffic congestion greatly increases the time it takes for emergency response units to reach the scene which can mean property damage or casualties. We found (Nagarjuna *et al*., 2020) a transportation system designed for emergency travel. The researchers plan to install RFID transmitters on all emergency vehicles such as ambulances, police cars, and fire trucks and install signal-reading devices on traffic lights. The system also uses the MQTT protocol to communicate on the Internet of Things based on ESP32 and Blynk. When the server receives the radio frequency that the vehicle is in emergency response, it will change the color of the signal lights on the path of the emergency vehicle in advance and arrange for police to keep the road clear.

Today, with the development of science and technology, governments of all countries have established their disaster prevention systems. Natural disasters once led to the destruction of human civilization and now we can often perceive disasters such as tsunami warnings, earthquake warnings, and hurricane warnings in advance. But there are few early warning systems for general emergencies. The researchers (Elvas *et al*., 2020) used data mining methods to extract knowledge from data of past events and apply it to the prediction of future events. Researchers follow the CRISP-DM method to use CIC data (fires, floods, industrial gas leaks, elevator failures, etc.) and translate it into knowledge.

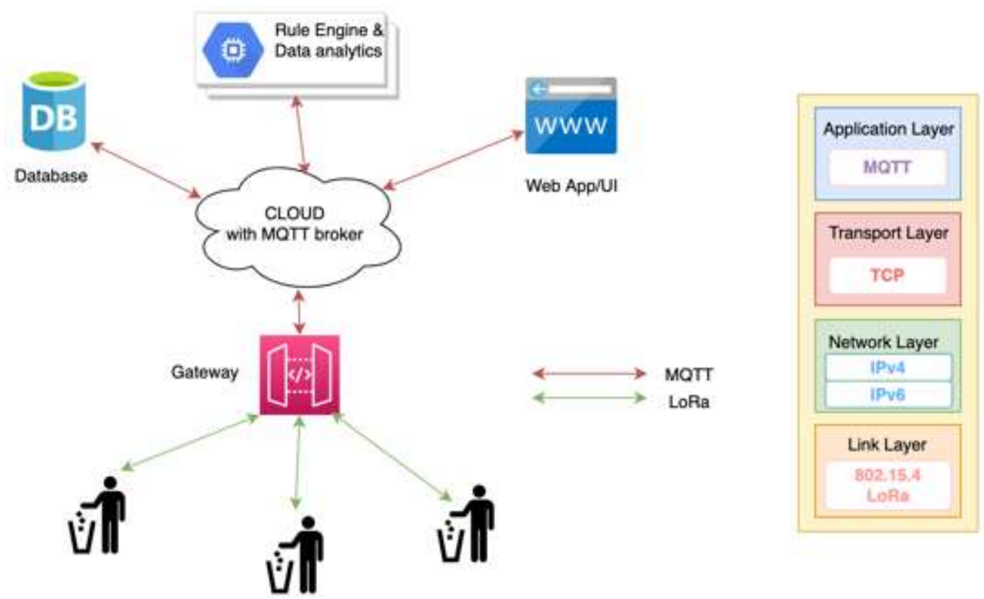

Fig. 6: Architecture of waste management system and protocol stack

DOI: 10.3844/jcssp.2022.396.414

Scientists in this research mainly use logistic regression, KNN, SVM, random forest algorithm, and decision tree to classify all emergency events. After 20 iterations, the most suitable prediction model is selected and the prediction results are intuitively displayed on the map (Fig. 7) according to data visualization technology.

In (Yang *et al*., 2017), the researchers applied big data to realize the intelligence of the whole process of urban emergency response, including prevention, protection, mitigation, response, and even post-disaster recovery, to improve the overall emergency response service level of the city. The system uses data generated by social media, smart sensors, and the Internet, mines GPS locations, analyzes important places where people gather, and compares it with monitoring system data to model and predict potential hazards in advance near high-risk buildings and areas. Deploy emergency response units. The biggest feature of this system is that it can conduct emergency response modeling and prediction based on the event prediction results, such as calculating evacuation routes in advance and obtaining the optimal emergency response work plan through simulation. However, most of the currently mentioned methods for making emergency response services more intelligent rely heavily on data, networks, and communication. Once data transmission is interrupted or data is lost due to a cyber-attack, the intelligent emergency response service we have built will no longer work. Therefore, to prevent the emergency response service from being stopped due to interruption of data flow, we found that (Kotevska *et al*., 2017) proposed a dynamic network model to improve the resilience of the service to data loss. This technology can be applied to various systems in smart cities to reduce the risk of service termination due to data loss.

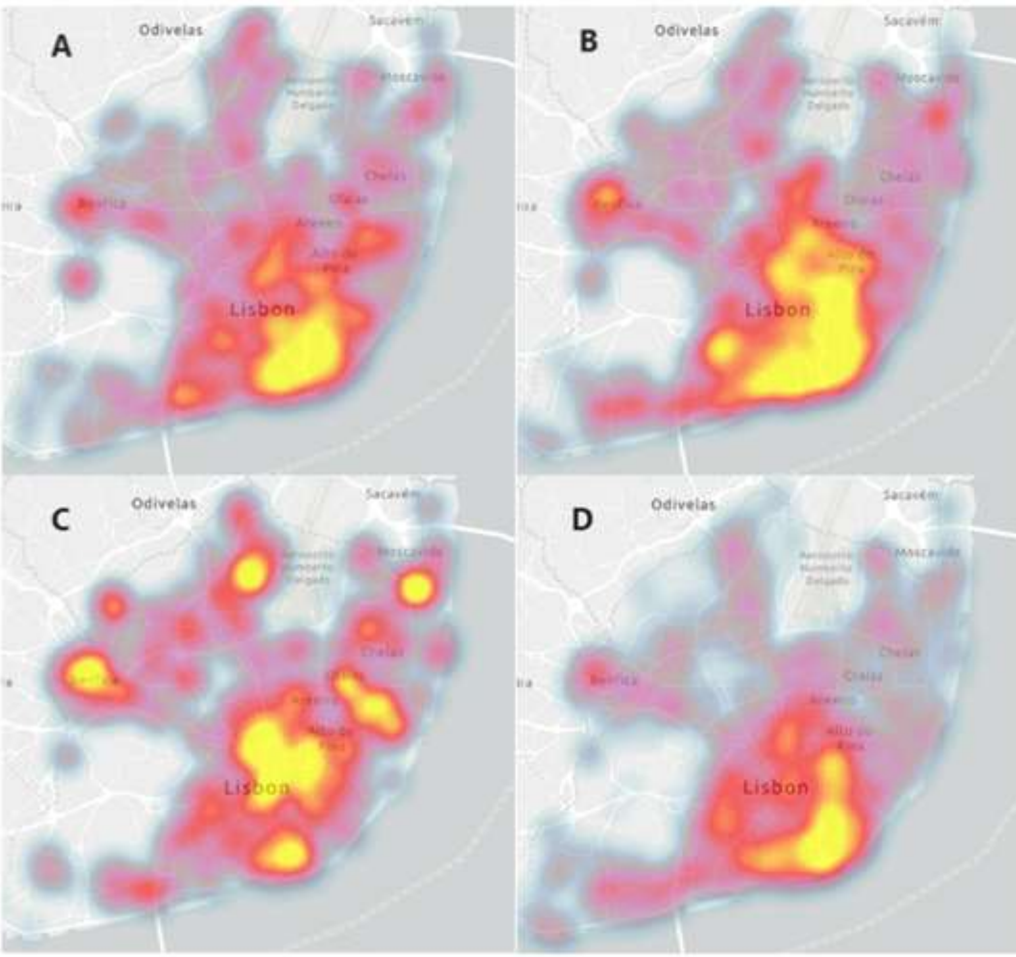

Fig. 7: Heatmap-Spatial Distribution of each type of occurrence group. The A subfigure shows fire (building fires), B subfigure shows industrial-technological (gas leaks and suspicious situations). fire (building fires), the C subfigure shows accidents (with equipment or with elevators) and the D subfigure shows infrastructure (floods and falls) (Elvas *et al*., 2020)

Smart City Planning

Urban development is inseparable from appropriate urban planning. In the past urban planning, professionals in the planning field and government Fig. 8a Point density and b Kernel Density Estimation (KDE) of check-ins distribution in the study area. Ali Haidery *et al*. (2020) decision-makers participated in the formulation of urban strategies. In the smart city concept, due to the support of sensors and the Internet of Things, urban planning will be more and more based on the data generated by citizens, truly achieving an organic integration of the city and human beings. To create smart cities, we need city planners who are willing to adopt data analytic frameworks. There are fourteen obstacles to the development of a smart city outlined in (Adil and Khan, 2021) which need to be addressed by city planners.

Economic construction is usually the primary consideration in urban planning. Appropriate urban function division can not only improve the happiness of citizens' life but also maximize the function of urban resources to generate the greatest economic value. We found several studies on data mining methods for classifying the way urban land and public space are used and thus more appropriately functionally zoning undeveloped land. In (Duan *et al*., 2020), scientists took the commercial services of rail transit stations as an example and used a Data Envelopment Analysis (DEA) model to evaluate the degree of coupling between the operational efficiency of rail transit stations and the relationship between land use. The higher the degree of coupling, the higher the mutual support between the development of surrounding land and the sustainable development of the station. This can help the planning of urban public transport lines, stations, and commercial land near stations. To study the utilization efficiency of urban public space more generally, (Lau *et al*., 2017) used IoT sensors to monitor public space. Researchers use WSN communication to achieve large-scale deployment of sensors, which monitor motion, noise, and other signals through sensors to determine the frequency of activities in public spaces. Liu *et al*. (2020) also conducted a similar study, but they used the data set provided by social software to analyze the data samples to describe the behavior patterns of millions of people using urban green spaces and according to the behavior patterns, parks or green spaces can be more appropriately arranged.

The upgrading and development of urban functions are inseparable from the citizens. Studying population mobility and urban population density can help policymakers formulate corresponding development strategies. In the context of COVID-19, data monitoring of population movements can increase the perception of the infectiousness of the virus. Further research incorporates human mobility data into epidemic risk prediction models based on the interaction of population mobility and epidemics (Wang *et al*., 2020). Specifically, taking Shanghai population density as the research object (Ali Haidery *et al*., 2020), the researchers also analyzed the spatial behavior characteristics of the population through Kernel Density Estimation (KDE) based on social media data (Fig. 8).

At the same time, urban civil air defense projects are often ignored by the public. Civil air defense engineering can ensure the basic living needs of residents in wartime or extreme disaster situations. In modern super cities with extremely high population density, civil air defense works and fire hydrants cannot meet the requirements stipulated by law. Laadan *et al*. (2020) The researchers compared the distribution data of urban air-raid shelters and fire hydrants with the construction standards of urban civil air defense engineering facilities and then used visualization technology to display the safety levels of different areas.

Another paper introduces a framework to unite the different strategic visions of a smart city under one common goal across different city sectors, allowing all of the recent growth in smart city technology solutions to be implemented in actual cities (Westraadt and Calitz, 2020). To be more efficient, a city should use Integrated City Management Platforms (ICMPs) to analyze data and identify how this data can be used across different sectors of a city interdependently. The study developed a quantitative tool to foster cross-sector collaborations which is meant to complement Key Performance Indicator frameworks and it focused specifically on crime management. They developed a multi-sector approach to crime management to demonstrate their strategy. To build a modeling framework, they propose four criteria: (1) determine the relationship of crime management with all relevant city sectors, (2) identify negative impacts of other sector activity on crime using a predictive model, and (3) handle large amounts of data and (4) accommodate complex interactions with city sectors and their underlying features. To demonstrate the proposed model, they ran it on New York City crime data. The study identified 4 separate “states” that were highly characteristic of crime areas: Collective Efficacy I, Commercial Land Use, Collective Efficacy II, and Mixed Land Use. Based on the information that the model produces, the proper governmental agencies can be identified to try and solve each characteristic relating to high crime levels. This example of using a model for identifying cross-sector indicators for certain behaviors (namely crime) and evaluation of said indicators is proposed. This model should foster cross-sector collaborations and give insights into how different aspects of a smart city can work together to find mutually beneficial solutions to problems using ICMPs. The authors believe that this strategy of identifying cross-sector indicators could in the future be applied to sectors such as energy and water usage.

DOI: 10.3844/jcssp.2022.396.414

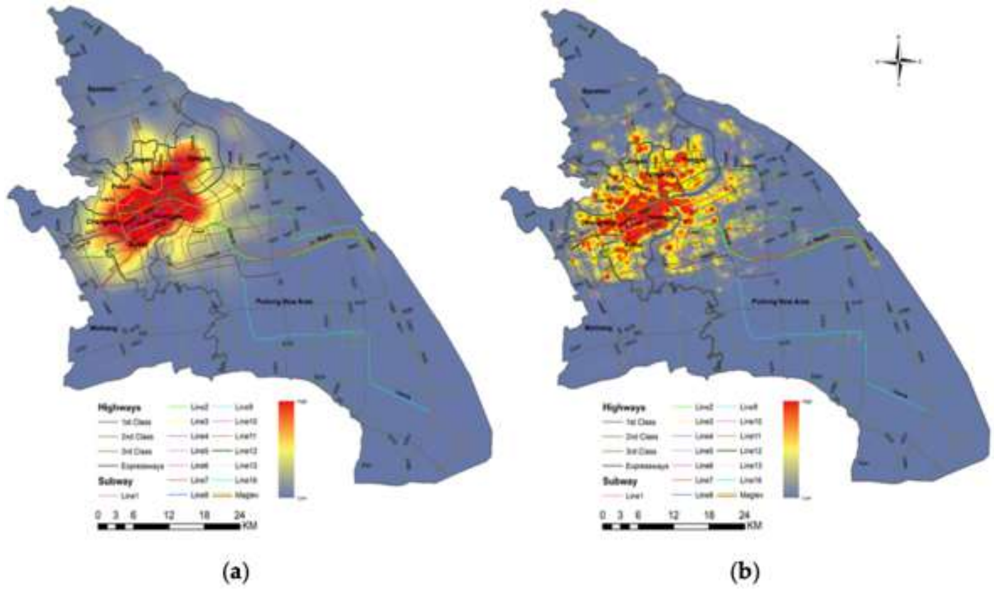

Fig. 8: (a) Point density and (b) Kernel Density Estimation (KDE) of check-ins distribution in the study area (Ali Haidery *et al*., 2020)

Smart Network and Privacy

Security is always a concern in a smart city, especially when personal data is being collected and shared with information systems. For obvious reasons, a smart city should not be given unadulterated access to the entirety of every resident's data, even though this would probably result in more accurate solutions across various sectors of the city, from transportation to energy usage. Privacy is a value that must be upheld when working with large amounts of data because a poorly designed network could fall prey to malicious attackers who could use personal data for harm rather than for welfare. It is in the highest interest of a smart city, then, to create solutions that are both secure from attacks and still useful to data mining techniques.

With the application of a large number of IoT devices such as sensors, smart wearable devices, and smart home appliances, the number of nodes connected to the network has increased greatly and the data transmitted on the network is unprecedentedly large. The encryption settings of the network access nodes of these many households are not taken seriously, which leads to device hijacking, data, and privacy leakage. In the past, many devices and applications did not respect users' privacy and data security. As the law improves, most countries prohibit the collection and processing of information unless the data subject provides informed and specific uniformity. Dhungana *et al*. (2015) introduced a privacy-preserving framework based on a set of Privacy-Preserving Agents (PPA), which are responsible for enforcing data access policies and controlling the data transformation process to protect data confidentiality and privacy. Each agent is independent of the other and is responsible for a specific domain. At the same time, the researchers also proposed three methods to improve data security: anonymization, data aggregation, and data perturbation and randomization. As the concept of data security continues to gain popularity, data protection technologies and concepts are also evolving. Mohamed *et al*. (2020) proposed four distinct applications for data-driven cybersecurity: (1) Interactive threat detection and mitigation applications, such as intrusion detection and prevention systems (IDPS); (2) prediction of possible attacks and risks; (3) Application of optimal configuration and management of security systems; (4) Application of systematic security planning for the future.

Specifically, (Khedr *et al*., 2019) provide a cluster-based scheme to mine sensor data. This approach does not require data transfer to cluster heads or base stations and enables maximum performance in a WSN environment. The researchers hope to determine the association rules of the global implicit database with the least exchange of information, also use data aggregation techniques and agents, and decompose the global computation into distributed computations to ensure the privacy and confidentiality of the data to the greatest extent. When it is known in advance that the sensor involves sensitive data, the data processing has to become more cautious. (Amma and Dhanaseelan, 2018) discusses how to protect sensitive data and methods for classifying sensitive data. The researchers used homomorphic encryption to encrypt the private data and then used the Naive Bayes algorithm to classify the data. To improve the performance of the data mining classifier, the researchers set up the computation to be done in the cloud, but they used homomorphic encryption locally before uploading the data to the cloud. In addition to improving data mining methods, it is equally important to design a rigorous and complete data security architecture. Hui *et al*. (2020) proposed a three-layer data security architecture for smart cities named "SafeCity". (1) Data security layer: only allow data to be located between real physical devices for data transmission and use symmetric encryption algorithms to prevent intrusion; (2) Data computing layer: Based on resource-intensive secure computing (3) Decision layer: Artificial intelligence-based decision agent. The most fundamental way to improve data security is to reduce data collection. However, many applications in our lives collect as much user data as possible. Li *et al*. (2015) studied the current situation of data over-collection and proposed a mobile cloud framework to eliminate the over-collection of data. Modern smartphones have many permissions to access user privacy such as accessing photo albums, tracking data, accessing contacts, accessing calendars, and tracking IMEI/UDID. These private access processes all increase the possibility of data leakage. Today, many mobile phone providers provide cloud storage services, but the security of cloud services is still not perfect. The researchers also believe that encrypting data before it is transmitted to the cloud can effectively improve security.

We found a case applied to smart grid intrusion detection. Subasi *et al*. (2018) Researchers designed an intrusion detection system (Fig. 9) that uses data mining techniques to detect attacks including illegal access to information in smart grids. The researchers found potentially intruding nodes in the intelligent network with poorly configured firewalls, network sector backdoors, and remote terminal units. The data classification methods, in this case, are ANN, KNN, SVM, CART, and RF. These classifiers learn and classify access types to detect illegal access and virus intrusion. The results show that all data mining methods have a performance of more than 95%, among which random forest has the highest accuracy.

Zhu *et al*. (2018) are written about security and privacy concerns related to the smart grid. As more and more users join the smart grid by implementing wireless smart meters in their homes, they could become vulnerable to certain kinds of cyberattacks, which are outlined in this study. The two main concerns when it comes to wireless communication networks are cyberattacks and privacy leakage. The reason why energy patterns are potentially important is that an adversary could analyze the energy usage data to discover personal activity habits or even approximate a daily schedule. Through various methods, this adversary could eavesdrop on data transmission, which would allow them to use big data mining algorithms to determine what appliances are being used. Using supervised or unsupervised machine learning models, one can determine when specific appliances are being used by analyzing when the energy consumption increases or decreases suddenly.

However, there are measures we can take to prevent adversaries from gaining this knowledge. They describe three strategies for defense, one for each data eavesdropping, traffic analysis, and information theory analysis. One strategy against data mining techniques involves a rechargeable battery that discharges and recharges at random intervals to shroud the relationship between the energy usage and the appliances. The battery could either be truly random, or it could be used as an offset to always keep the energy usage at an even level. Either way, it would be nearly impossible for an adversary to identify usage patterns if a rechargeable battery was being used in one of these ways.

Smart IoT Data and Smart Service and Application

As more and more of our devices have gained the capability to talk amongst one another and share information, we have seen the development of the Internet of Things (IoT). The IoT allows us to gain insights into human behavior that was never possible before and it provides an excellent avenue for data mining techniques. Since the dawn of the internet, more and more devices are being connected every year (Fig. 10) and this figure will only continue to grow. The wealth of knowledge extrapolated from these connected devices allows us to incorporate data mining using Artificial Intelligence (Priyanka and Thangavel, 2020). Because of the massive amounts of data that the IoT produces, it is very important to have analytic and sustainable methods to make sense of all this data.

Sodhro *et al*. (2019) Sensors are critical to the implementation of a smart city because they are the means through which all of the data must be collected. This study discusses the critical role of sensors in a smart city, but also brings to light how much energy consumption they create in the process. To help solve this problem, the authors propose a Hybrid Adaptive Bandwidth and Power Algorithm (HABPA) and Delay-tolerant Streaming Algorithm (DSA). The authors also offer a novel framework for IoT for smart devices in a smart city, arguing that the power and battery of IoT devices should be optimized with minimum delay through buffers in data transmission.

The paper found video transmission to be a large cause of energy usage that can be optimized. They propose an architecture for smart cities in which power drain and battery life are taken into account and calculated at various stages throughout the data pipeline process. This is a relatively novel idea because most studies do not consider the amount of energy consumption in smart devices. Among other things, they found that a small buffer size drains power and battery life in IoT devices. Both proposed algorithms perform better than the baseline algorithms but have some limitations. The implementation of these algorithms could help a city make significant progress toward greener energy while still making progress in the other domains. Approaches like these truly define what we strive for in a smart city.

In addition to resource management on a city-wide level, it is also important to consider how a connected network of IoT devices might be used to create smart buildings. Daissaoui *et al*. (2020) lays out the framework for how a smart city would go about the process of turning its buildings into smart buildings. They propose that a combination of sensors, networks, big data, and analytics should be used in buildings to create an application that lets a user manage the facilities and energy, track resources and occupants, and enhance comfort for those in the building.

Data mining technology can also analyze citizens' attitudes towards daily life status or policies. Kharlamov *et al*. (2021) The researchers found that due to the COVID-19 pandemic, there are differences in how well people adapt to IT technology. Digital services in smart cities have received mixed reviews. They used a multimodal approach, and neural network technology to perform textual analysis, sentiment analysis, and lexical association of words on social media. And using conditional random field methods such as the maximum entropy Markov model, the results show that the most popular are online shopping, digital cinema, and telemedicine. However, the most negative impressions are distance learning, digital bank passes, webcams, etc. A similar study also occurred (Kontokosta and Hong, 2021), where researchers used data mining techniques to identify differences in residents' interactions with the government, based on the Missouri 311 report. And under- and over-reporting communities are categorized based on complaining behavior. The results showed that low-income households were primarily classified as underreporting, while high-income households were classified as over reporting. Interestingly, the measures taken by the government for low-income neighborhoods are far less than for high-income neighborhoods.

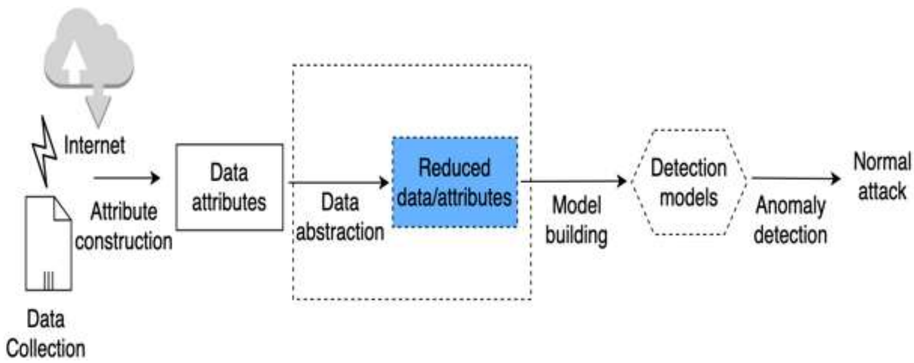

Fig. 9: Intrusion detection steps

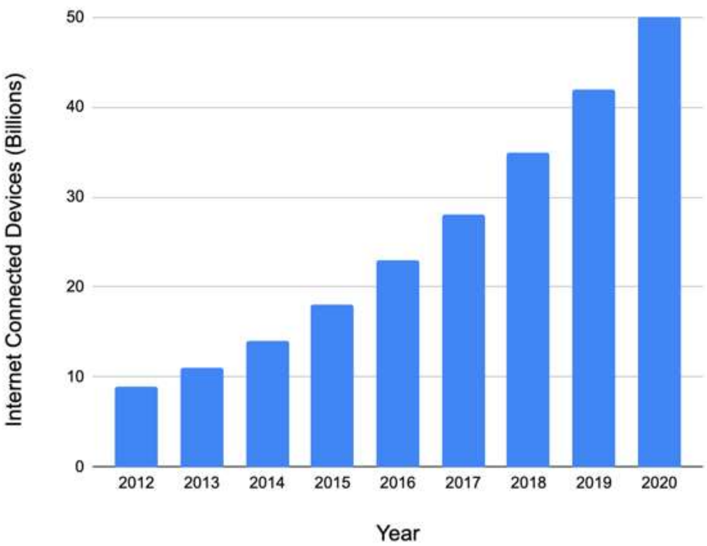

Fig. 10: Increase in internet-connected devices by the year

Table 1: Comparison of existing literature

Ref.	Title	Results/Methodology	Limitations/Future Work
Al Nuaimi *et al*. (2015)	Proactive big data analysis for traffic accident prediction	Traffic risk is assessed and shared with drivers. Read traffic volume was reduced by 20% and the average speed was reduced by 40% in dangerous conditions.	More work should be done to implement these strategies in the real world and judge their effectiveness at predicting accidents.
Finogeev *et al*. (2020)	Traffic video-based intelligent traffic control system for smart cities using modified ant colony optimizer	A Support Vector Machine estimates traffic on road segments, then an Ant Colony Optimizer Algorithm determines the optimal path for vehicles.	Would be much more accurate with vehicle-to-vehicle communication.
Liu and Wang's (2019)	Design of traffic emergency Response system based on internet of things and data mining in Emergencies	Proposes that cars should be given RFID bracelets for emergencies to make traffic analysis easier. Found that a Bayesian network model would perform better than others in emergencies.	More accurate data collection is needed, as well as applying these techniques for other types of disasters.
Yousef *et al*. (2019)	Intelligent traffic light scheduling technique using calendar-based history information	Using year-round traffic data, light management algorithms are improved to predicts traffic levels before they happen.	Performs better in busy traffic times and worse in normal traffic. Only applied in a simulation.
Gera *et al*. (2021)	IoT based automated health care monitoring system for smart city	Cloud storage is utilized with smart biosensors to store biological information, promoting telemedicine in the future.	This smart health monitoring system can add artificial intelligence to facilitate automated functions.
Pei *et al*. (2020)	Minimal green energy consumption and workload management for data centers on smart city platforms	A hybrid genetic approach (SSPD-HG) is used to control the energy usage of data centers, resulting in increased energy efficiency and load management.	Problems with overlapping priorities in different operating environments require a redesign of the model's decision-making mechanism.
VE *et al*. (2021)	Efficient energy consumption prediction model for a data analytic-enable industry building in a smart model. a city	Smart meters track the energy consumption of equipment, which is then fed into a predictive Results show that the SVM model has with the highest stability and prediction accuracy.	This method is only for the steel industry and the model needs to been optimized to apply to more industries in the future.
Alanezi *et al*. (2021)	Automated Residential Energy Audits Using a Smart WiFi Thermostat-Enabled Data Mining Approach	A machine learning model was trained to predict residential energy efficiency by by installing a smart WIFI thermostat. Homes' with GBM density prediction model had the highest energy thermal efficiency.	The light transmittance of the architectural glass also has an impact on household energy consumption and future research should consider the dynamic the input of solar radiation.
Golpîra and Bahramara (2020)	Internet-of-things-based optimal smart city energy management considering shiftable loads and energy storage	A Smart Grid which utilizes an edge computing framework is proposed, which reduces the total operating cost of the urban power grid by about 14.28%.	Research also needs to account for smart grid node failures and sensor data noise.
Oralhan *et al*. (2017)	Smart city application: Internet of Things (IoT) technologies based smart waste collection using data mining approach and ant colony optimization	Garbage collection is optimized using an Ant Colony Algorithm and smart sensors are placed in waste bins to improve efficiency. Waste collection costs were improved by 30%.	Regional population levels can be analyzed and future planning can be done through a waste receptacles and collection data.
Bharadwaj *et al*. (2016)	IoT based solid waste management system: A conceptual approach with an architectural solution as a smart city application	LoRa is used to achieve long-distance data transmission with low power consumption, and waste collection routes are optimized using Google API.	A data transport layer security protocol can be set between the terminal device and the gateway to improve communication security. And the management system is integrated into the biogas monitoring system.
Nagarjuna *et al*. (2020)	IoT Enabled Smart Traffic System for Public and Emergency Mobility in Smart City	RFID transmitters are used in emergency vehicles through IoT to improve emergency response times in traffic.	Apply artificial intelligence technology to the intelligent traffic management system to further, improve the level of urban intelligence.
Elvas *et al*. (2020)	Data-Driven Approach for Incident Management in a Smart City	An early warning system for natural emergencies is presented using different prediction models, which are then mapped onto a heat map of a region.	The collected data is not enough to support the prediction accuracy. More people and devices need to be connected to the system to obtain more prediction data and improve the prediction accuracy.
Yang *et al*. (2017)	Using big data to enhance crisis response and disaster resilience the smart city	A prediction model predicts hazards in advance based on data generated by smart sensors setting around the buildings.	Apply big data technologies and tools to solve specific types of natural disasters or large-scale accidents.
Duan *et al*. (2020)	Operating Efficiency-Based Data Mining on Intensive Land Use Smart City	The density inside the transit station zone should not be lower than 7.5 km^2 and bus line density should above 55 km^2.	This study only focuses on commercial land and more attention should be paid to the use of other land types around rail0 transit in the future.
Liu *et al*. (2020)	Categorization of Green Spaces for a Sustainable Environment and Smart City Architecture by Utilizing Big Data	The neighborhood green parks are much more crowded than other green spaces in Shanghai. Also, the number of check-ins is much higher in summer and spring.	LBSN data has limitations such as low sample size frequency, location class bias, etc. Therefore, LBSN data can only be used as a supplement to traditional data sources.
Laadan *et al*. (2020)	Using Data Mining for Infrastructure and Safety Violations Discovery in Cities	Visually display areas that violate fire and safety regulations on a map geographic information data projection and analyze violations in different parts of the city.	In the future, automated systems can be built to automatically plan build new fire safety facilities based on data projection results.
Westraadt and Calitz (2020)	A Modelling Framework for Integrated Smart City Planning and Management	Data from across all sectors of a city is used to create a unified model. Crime management was shown as an example and the model discovered four characteristics that predict high crime within a city.	In the future, this technique should be applied to energy and water usage.
Dhungana *et al*. (2015)	Aspern smart ICT: Data analytics and privacy challenges in a smart city	Under strict privacy and security regulations, new data security technologies can greatly improve the security of personal privacy data in smart cities.	Explore new privacy and access control solutions backed by external security audits.
Khedr *et al*. (2019)	Privacy-Preserving Data Mining Approach for IoT based WSN in Smart City	The scheme utilizes the computing power for each sensor node to performing distributed computing, reducing the amount of data transmission and reducing energy consumption and data leakage.	The security of data transmission will be tested in a wider range of WSN environments in the future.
Subasi *et al*. (2018)	Intrusion Detection in Smart Grid Using Data Mining Techniques	After 10-fold cross-validation, the random forest detection model has the highest detection accuracy of 98.78%.	In the future, this detection technology can be installed at every terminal of the smart grid, such as household electricity meters and community transformers.
Zhu *et al*. (2018)	Big Data Mining of Users' Energy Consumption Patterns in the Wireless Smart Grid	Three strategies are given to prevent adversaries from mining grid data, protecting privacy. A rechargeable battery system that fluctuates randomly is proposed.	Effective privacy attacks can be researched to build-up defenses, as well as the extension of these principles to other resources such as water or coal.
Priyanka and Thangavel (2020)	Towards an Optimal Resource Management for IoT-based Green and Sustainable Smart Cities	Two algorithms (HABPA and DSA) were used to analyze energy usage in cities. Battery life of IoT devices are diminished when the buffer size of transmitted data is decreased.	Further work can be done in optimizing buffer size and battery drain while still collecting useful and timely data.
Kontokosta and Hong (2021)	How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions	Reports from Missouri 311 complaints are analyzed about the income levels of the neighborhood in which they originated and classified as underreporting or overreporting. Complaints from high-income neighborhoods resulted in more governmental action.	This analysis can be applied other 311 reporting data sources in different cities to better identify discrepancies between facts and reports.

Results and Discussion

The table above (Table 1) identifies the strongest and most applicable papers from each domain, compiled together. This is useful because it allows us to better synthesize and analyze the findings of the Survey of Current Research section of this study. Many of the models presented have yet to be implemented in the real world, which provides us with a starting point when we look forward to future research. Smart city researchers can examine the limitations of these main papers to provide real-world solutions to these problems. When we can view all of this research together, we will begin to visualize and realize cross-sector innovations in data mining in a smart city context.

Conclusion and Future Work

Smart cities are the way of the future. With the increase in urban population worldwide, there also comes an increase in negative consequences such as congestion, waste, crime, lack of privacy, and more. In addition to using solutions developed in the past, we need to also look forward to novel solutions. After all, modern problems require modern solutions. Sensors and interconnected devices are growing at a rate that mirrors the exponential population growth, which lends itself to data mining. We must analyze data mining techniques to promote the well-being of humans, not for profit or strict governmental control. We identified seven critical domains of a smart city: Smart Transportation, Smart Healthcare, Smart Energy, Smart City Utilities, Smart City Planning, Smart Networks, and Privacy and Smart IoT applications. For smart cities to become a reality, we must continue our research in each of these areas.

In the future, we recommend that more research be done regarding public transportation optimization. With Smart Transportation, if we only try to make car networks more efficient, we may find that we are merely treating a symptom rather than a deeper societal problem. Another essential area of future research is Smart Networks and Privacy. This is because data privacy must be applied to every single domain of a smart city. For data mining solutions to be implemented in the real world, people need to be convinced that their data is not at risk of governmental abuse. Research should be done to protect private data, especially data taken from cameras and sensors in a smart city. We hope that one day we will see our cities fully transformed into efficient, sustainable, safe, and healthy spaces for all who abide within.

Acknowledgment

We thank office of research and sponsored programs for supporting this faculty-student collaboration research project at University of Wisconsin Eau Claire.

Author’s Contributions

Rushit Dave: Analysis, Review, writing.

Nuo Xu and Simon Arneberg: Analysis, writing, narrating.

Naeem Seliya and Mounika Vanamala: Analysis Review.

Ethics

The authors declare that there exist no known competing financial interests that could have influenced the work in this study. The authors approve the publication of manuscript.

References

Adil, M., & Khan, M. K. (2021). Emerging iot applications in sustainable smart cities for COVID-19: Network security and data preservation challenges with future directions. Sustainable Cities and Society, 75, 103311. doi.org/10.1016/j.scs.2021.103311

Akbarpour, N., Salehi-Amiri, A., Hajiaghaei-Keshteli, M., & Oliva, D. (2021). An innovative waste management system in a smart city under stochastic optimization using vehicle routing problems. Soft Computing, 25(8), 6707-6727. doi.org/10.1007/s00500-021-05669-6

Akca, T., Sahingoz, O. K., Kocyigit, E., & Tozal, M. (2020). Intelligent Ambulance Management System in Smart Cities. 2020 International Conference on Electrical Engineering (ICEE), 1–7. doi.org/10.1109/ICEE49691.2020.9249959

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. Journal of Internet Services and Applications, 6(1), 1-15. doi.org/10.1186/s13174-015-0041-5

Ali Haidery, S., Ullah, H., Khan, N. U., Fatima, K., Rizvi, S. S., & Kwon, S. J. (2020). Role of big data in the development of smart city by analyzing the density of residents in Shanghai. Electronics, 9(5), 837. doi.org/10.3390/electronics9050837

Amma, N. N., & Dhanaseelan, F. R. (2018, October). Privacy preserving data mining classifier for smart city applications. 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 645-648. IEEE. doi.org/10.1109/CESYS.2018.8724081

Alanezi, A., Hallinan, K., & Huang, K. (2021). Automated Residential Energy Audits Using a Smart WiFi Thermostat-Enabled Data Mining Approach. Energies. 14(9), 2500. doi.org/10.3390/en14092500

Babar, M., Khattak, A. S., Jan, M. A., & Tariq, M. U. (2021). Energy aware smart city management system using data analytics and Internet of Things. Sustainable Energy Technologies and Assessments, 44, 100992. doi.org/10.1016/j.seta.2021.100992

Belhadi, A., Djenouri, Y., Srivastava, G., Djenouri, D., Lin, J. C. W., & Fortino, G. (2021). Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection. Information Fusion, 65, 13-20. doi.org/10.1016/j.inffus.2020.08.003

Bharadwaj, A. S., Rego, R., & Chowdhury, A. (2016, December). IoT based solid waste management system: A conceptual approach with an architectural solution as a smart city application. 2016 IEEE annual India conference (INDICON), 1-6. IEEE. doi.org/10.1109/INDICON.2016.7839147

Daissaoui, A., Boulmakoul, A., Karim, L., & Lbath, A. (2020). IoT and Big Data Analytics for Smart Buildings: A Survey. Journal of Ubiquitous Systems and Pervasive Networks. 13. 27-34. doi.org/10.5383/JUSPN.13.01.004

Dhungana, D., Engelbrecht, G., Parreira, J. X., Schuster, A., & Valerio, D. (2015, December). Aspern smart ICT: Data analytics and privacy challenges in a smart city. 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), 447-452. IEEE. doi.org/10.1109/WF-IoT.2015.7389096

Duan, Y. Q., Fan, X. Y., Liu, J. C., & Hou, Q. H. (2020). Operating efficiency-based data mining on intensive land use in smart city. IEEE Access, 8, 17253-17262. doi.org/10.1109/ACCESS.2020.2967437

Elvas, L. B., Marreiros, C. F., Dinis, J. M., Pereira, M. C., Martins, A. L., & Ferreira, J. C. (2020). Data-driven approach for incident management in a smart city. Applied Sciences, 10(22), 8281. doi.org/10.3390/app10228281

Enler, E., Pentek, I., & Adamko, A. (2020, September). Healthcare Framework for Smarter Cities with bio-sensory data. 2020 11th IEEE International Conference on Cognitive in communications (CogInfoCom), 000337-000342. IEEE. https://ieeexplore.ieee.org/abstract/document/9237838/

Finogeev, A., Deev, M., Finogeev, A., & Kolesnikoff, I. (2020, November). Proactive Big Data Analysis for Traffic Accident Prediction. 2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA), 1-9. IEEE. doi.org/10.1109/CITISIA50690.2020.9371796

Gera, S., Mridul, M., & Sharma, S. (2021, April). IoT Based Automated Health Care Monitoring System for Smart City. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 364-368. IEEE. doi.org/10.1109/ICCMC51019.2021.9418487.

Golpîra, H., & Bahramara, S. (2020). Internet-of-things-based optimal smart city energy management considering shiftable loads and energy storage. Journal of Cleaner Production, 264, 121620. doi.org/10.1016/j.jclepro.2020.121620

Han, J., Song, W., Gozho, A., Sung, Y., Ji, S., Song, L., ... & Zhang, Q. (2020). Lora-based smart IoT application for smart city: An example of human posture detection. Wireless Communications and Mobile Computing, 2020. doi.org/10.1155/2020/8822555

Honarvar, A. R., & Sami, A. (2019). Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures. Big data research, 17, 56-65. doi.org/10.1016/j.bdr.2018.05.006

Hui, Z., Babar, M., Tariq, M. U., Ahmad Jan, M., Menon, V., & Li, X. (2020). Safe City: Toward Safe and Secured Data Management Design for IoT-Enabled Smart City Planning. IEEE Access. 1-1. doi.org/10.1109/ACCESS.2020.3014622

Kharlamov, A. A., Raskhodchikov, A. N., & Pilgun, M. (2021). Smart city data sensing during COVID-19: Public reaction to accelerating digital transformation. Sensors, 21(12), 3965. doi.org/10.3390/s21123965

Khedr, A. M., Osamy, W., Salim, A., & Salem, A. A. (2019). Privacy preserving data mining approach for IoT based WSN in smart city. International Journal of Advanced Computer Science and Applications, 10(8), 555-563. doi.org/10.14569/IJACSA.2019.0100873

Kontokosta, C. E., & Hong, B. (2021). Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. Sustainable Cities and Society, 64, 102503. doi.org/10.1016/j.scs.2020.102503

Kotevska, O., Kusne, A. G., Samarov, D. V., Lbath, A., & Battou, A. (2017). Dynamic network model for smart city data-loss resilience case study: City-to-city network for crime analytics. IEEE Access, 5, 20524-20535. doi.org/10.1109/ACCESS.2017.2757841

Laadan, D., Arviv, E., & Fire, M. (2020, July). Using Data Mining for Infrastructure and Safety Violations Discovery in Cities. 2020 IEEE International Smart Cities Conference (ISC2), 1-4. IEEE. doi.org/10.1109/ISC251055.2020.9239035

Lau, B. P. L., Wijerathne, N., Ng, B. K. K., & Yuen, C. (2017). Sensor fusion for public space utilization monitoring in a smart city. IEEE Internet of Things Journal, 5(2), 473-481. doi.org/10.1109/JIOT.2017.2748987

Li, Y., Dai, W., Ming, Z., & Qiu, M. (2015). Privacy protection for preventing data over-collection in smart city. IEEE Transactions on Computers, 65(5), 1339-1350. doi.org/10.1109/TC.2015.2470247

Liu, Q., Ullah, H., Wan, W., Peng, Z., Hou, L., Rizvi, S. S., ... & Muzahid, A. A. M. (2020). Categorization of green spaces for a sustainable environment and smart city architecture by utilizing big data. Electronics, 9(6), 1028. doi.org/10.3390/electronics9061028

Liu, Z., & Wang, C. (2019). Design of traffic emergency response system based on internet of things and data mining in emergencies. IEEE Access, 7, 113950-113962. doi.org/10.1109/ACCESS.2019.2934979

Lokuliyana, S., Jayakody, A., Dabarera, G. S. B., Ranaweera, R. K. R., Perera, P. G. D. M., & Panangala, P. A. D. V. R. (2018, August). Location based garbage management system with IoT for smart city. 2018 13th International Conference on Computer Science & Education (ICCSE), 1-5. IEEE. doi.org/10.1109/ICCSE.2018.8468682

Mohamed, N., Al-Jaroodi, J., & Jawhar, I. (2020, July). Opportunities and challenges of data-driven cybersecurity for smart cities. 2020 IEEE Systems Security Symposium (SSS), 1-7. IEEE. doi.org/10.1109/SSS47320.2020.9174388

Nagarjuna, G. R., Shashidhar, R., Puneeth, S. B., & Arunakumari, B. N. (2020, October). IoT Enabled Smart Traffic System for Public and Emergency Mobility in Smart City. 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) 53-59. IEEE. doi.org/10.1109/I-SMAC49090.2020.9243489

Oralhan, Z., Oralhan, B., & Yiğit, Y. (2017). Smart city application: Internet of things (IoT) technologies based smart waste collection using data mining approach and ant colony optimization. Internet Things, 14(4), 5. https://iajit.org/PDF/%20Vol%2014,%20No.%204/14969.pdf

Pal, D., Triyason, T., & Padungweang, P. (2018). Big data in smart-cities: Current research and challenges. Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 6(4), 351-360. doi.org/10.11591/ijeei.v6i4.543

Pei, P., Huo, Z., Martínez, O. S., & Crespo, R. G. (2020). Minimal green energy consumption and workload management for data centers on smart city platforms. Sustainability, 12(8), 3140. doi.org/10.3390/su12083140

Pfortmueller, C. A., Lindner, G., & Exadaktylos, A. K. (2014). Reducing fall risk in the elderly: Risk factors and fall prevention, a systematic review. Minerva Med, 105(4), 275-81.

Priyanka, E. B., & Thangavel, S. (2020). Influence of Internet of Things (IoT) In Association of Data Mining Towards the Development Smart Cities-A Review Analysis. Journal of Engineering Science & Technology Review, 13(4). doi.org/10.25103/jestr.134.01

Putra, A. S., & Warnars, H. L. H. S. (2018, September). Intelligent Traffic Monitoring System (ITMS) for Smart City Based on IoT Monitoring. In 2018 Indonesian Association for Pattern Recognition International Conference (INAPR) (pp. 161-165). IEEE. doi.org/10.1109/INAPR.2018.8626855

Ragavan, K., Venkatalakshmi, K., & Vijayalakshmi, K. (2021). Traffic video-based intelligent traffic control system for smart cities using modified ant colony optimizer. Computational Intelligence, 37(1), 538-558. doi.org/10.1111/coin.12424

Rahman, M. A., Hossain, M. S., Showail, A. J., Alrajeh, N. A., & Alhamid, M. F. (2021). A secure, private and explainable IoHT framework to support sustainable health monitoring in a smart city. Sustainable Cities and Society, 72, 103083. doi.org/10.1016/j.scs.2021.103083

Rayan, Z., Alfonse, M., & Salem, A. B. M. (2019). Machine learning approaches in smart health. Procedia Computer Science, 154, 361-368. doi.org/10.1016/j.procs.2019.06.052

Sari Aslam, N., Ibrahim, M. R., Cheng, T., Chen, H., & Zhang, Y. (2021). ActivityNET: Neural networks to predict public transport trip purposes from individual smart card data and POIs. Geo-spatial Information Science, 24(4), 711-721. doi.org/10.1080/10095020.2021.1985943

Sodhro, A. H., Pirbhulal, S., Luo, Z., & De Albuquerque, V. H. C. (2019). Towards an optimal resource management for IoT based Green and sustainable smart cities. Journal of Cleaner Production, 220, 1167-1179. doi.org/10.1016/j.jclepro.2019.01.188

Subasi, A., Al-Marwani, K., Alghamdi, R., Kwairanga, A., Qaisar, S. M., Al-Nory, M., & Rambo, K. A. (2018, April). Intrusion detection in smart grid using data mining techniques. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE. doi.org/10.1109/NCG.2018.8593124

VE, S., Shin, C., & Cho, Y. (2021). Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city. Building Research & Information, 49(1), 127-143. doi.org/10.1080/09613218.2020.1809983

Wang, A., Zhang, A., Chan, E. H., Shi, W., Zhou, X., &, Liu, Z. (2020). A review of human mobility research based on big data and its implication for smart city development. ISPRS International Journal of Geo-Information, 10(1), 13. doi.org/10.3390/ijgi10010013

Westraadt, L., & Calitz, A. (2020). A modelling framework for integrated smart city planning and management. Sustainable Cities and Society, 63, 102444. doi.org/10.1016/j.scs.2020.102444

Wray, A., Olstad, D. L., & Minaker, L. M. (2018). Smart prevention: A new approach to primary and secondary cancer prevention in smart and connected communities. Cities, 79, 53-69. doi.org/10.1109/ICSENS.2016.7808730

Wu, Y., Zhang, W., Shen, J., Mo, Z., & Peng, Y. (2018). Smart city with Chinese characteristics against the background of big data: Idea, action and risk. Journal of Cleaner Production, 173, 60-66. doi.org/10.1016/j.jclepro.2017.01.047

Yang, C., Su, G., & Chen, J. (2017, March). Using big data to enhance crisis response and disaster resilience for a smart city. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 504-507. IEEE. doi.org/10.1109/ICBDA.2017.8078684

Yousef, K. M. A., Shatnawi, A., & Latayfeh, M. (2019). Intelligent traffic light scheduling technique using calendar-based history information. Future Generation Computer Systems, 91, 124-135. doi.org/10.1016/j.future.2018.08.037

Zaree, T., & Honarvar, A. R. (2018). Improvement of air pollution prediction in a smart city and its correlation with weather conditions using metrological big data. Turkish Journal of Electrical Engineering & Computer Sciences, 26(3), 1302-1313. https://journals.tubitak.gov.tr/elektrik/abstract.htm?id=22717

Zhu, L., Li, M., Zhang, Z., Du, X., & Guizani, M. (2018). Big data mining of users' energy consumption patterns in the wireless smart grid. IEEE Wireless Communications, 25(1), 84-89. doi.org/10.1109/MWC.2018.1700157