Smart City Data Analysis

Mama Nsangou Mouchili
Concordia University
puter Science and Software Engineering Departm
55 Maisonneuve Blvd. W., Montreal, QC H3G 1M
+15145463065
mnsangou.mouchili@gmail.com

Shadi Aljawarneh
Software Engineering Department
JUST, Jordan
saaljawarneh@just.edu.jo

Wette Tchouati
Technology Specialist Data Science, System & Technology
BU IT and Cloud Products, Ericsson Canada
Constant.wette.tchouati@ericss@on.com

ABSTRACT

Smart City is one of the vital issues in the next coming years as it is estimated that more number of people will be migrating towards city and by 2040 cities is populated by 70% of the world's population. This will give raise to the city management problems like traffic congestion, parking queues, capacity planning and continuous depreciation in energy. Hence there is a need to understand the upcoming issues and find an efficient way to resolve them by resource planning.

The evolution of city to Smart City demand information and communications technology (ICT) to gather data. The data is collected based on city infrastructure and conduct of people in the city. This big data is analyzed to make knowledgeable decisions to give quality of life to the citizens. It is referred as the evolution of intelligent city to the next generation city of information technology. ICT provides real time processing of data to gather useful information. Based on this information visual applications can be developed to have a more farsighted view of the Smart City. Online banking, online shopping and remote video communication are some of the applications of Smart City.

As one of the concepts to improve city structure and its facilities, Smart City is vital in city management and development and uses information and communication technology (ICT) for city evolution and fast growth. These cities are planned to enhance more business and industrial growth thus increasing in economic development and raising living standard of the inhabitants [1]. It offers various facilities to the city like better transportation in the city, more job opportunities, less travelling time, healthy environment, better healthcare, less crime etc.

Smart City is a concept which works with sensors and data analytics to enhance the standard and quality of life in the cities. Smart City deals with the challenges like traffic, parking, capacity planning, energy etc. and resolves them to provide better city facilities to their citizens [2]. ICT is becoming widespread in the cities and is being used as a tool to deal with Smart City application domains. Smart City is using big data to generate intelligent information systems which support decision making capabilities. With effective data sources and data analytics tool we can design high end services for the citizens in urban cities. Big data has changed the way we store data and process the data efficiently. Now this crucial data is not only used to get the historic trends information but also used to predict the future. It is significant in planning the future of the cities. Data is analyzed to develop action plans from the predicted results.

In this paper, various datasets for traffic management and parking garage management have been tested and evaluated by using data analytics techniques in order to find a pattern in a historical information. Based on this research this information is further analyzed by machine learning tools and algorithms to predict the future behavior and help on taking decision. Machine learning and Analytics tools for Data Analytics from Hadoop and Elasticsearch environments as well as Mining algorithms from Python library are used for data prediction.

This paper gives us the details of some domain applications of Smart City listed as below:

• **Smart City Traffic Management**

• **Smart City Parking Management.**

DATA '18, October 1–2, 2018, Madrid, Spain © 2018 Association for Computing Machinery. ACM ISBN 978-1-4503-6536-9/18/10…$15.00
https://doi.org/10.1145/3279996.3280029

Keywords

Smart City, Data Analytics, Big Data, Random Forest, Mining Algorithms, Insights, city management, traffic congestion, capacity planning, traffic management, parking management.

1. INTRODUCTION

Designing and building an effective transport network is one of the biggest challenges facing cities. As the urban population grows, so the city's public and private transport systems become strained, and even small issues and incidents can adversely affect a whole network that is already operating close to its limits. Congestion and delays on transport systems can cost a city many millions of dollars in lost productivity and affect the quality of life of citizens as they spend more time travelling to their destination. The problems caused by congestion are immediately obvious – time wasted, increased pollution and safety concerns, impacting economic development and souring citizens' relationships with their local government. Cities need to address fundamental issues, such as the time drivers spend looking for parking places, improving road safety, ensuring public transport is routed effectively and giving citizens the ability to use multiple means of mobility to reach their destination.

In the past, public transport has operated on fixed routes, but the emergence of new 'mobility-as-a-service' business cases open up transport networks to become much more flexible in how citizens pay for, and use, cities' transport services. The integration of different transit mechanisms, such as bus networks and bike sharing, becomes easier to achieve, freeing up the transport planner to concentrate on service quality rather than service support. The flexibility enabled by connected services can improve the quality and breadth of transport provision. These improvements can enhance citizens' quality of life by reducing air and noise pollution, attracting new business investment or encouraging a healthier lifestyle through more walking and cycling. Therefore, developing connected transport services tends to be a core part of any smart city's objectives. [4]

ICT is becoming widespread in the cities and is being used as a tool to deal with Smart City application domains. Smart City is using sensors and big data to generate intelligent information systems which support decision making capabilities. With effective data sources and data analytics tool we can design high end services for the citizens in urban cities. Big data has changed the way we store data and process the data efficiently. Now this crucial data is not only used to get the historic trends information but also used to predict the future. It is significant in planning the future of the cities. Data is analyzed to develop action plans from the predicted results [5-7].

In this paper, we are using various datasets for traffic management and parking garage management. These datasets are used for data analytics in attempt to find potentially an approachable pattern in a historical information. Based on this research this information is further analyzed by machine learning tools and algorithms to predict the future behavior and help on taking decision. Machine learning and Analytics tools for Data Analytics from Hadoop and Elasticsearch environments as well as Mining algorithms from Python library are used for data prediction.

This paper gives us the details of some domain applications of Smart City listed as below:

• **Smart City Traffic Management.** Because of the growth of the number of vehicles, there is a need to manage traffic flow in the city to reduce traffic congestion and other issues. Smart City traffic management collects and analyzes the data based on traffic speed, density, speed limits, timestamp, location etc. Based on this information city can be better planned having balance mobility, safety and reducing environmental issues like fuel emissions.

• **Smart City Parking Management.** Parking is a challenging issue in cities, with average driver spending 6 to 14 minutes, in searching for parking, hence 30 percent of all traffic congestion is caused by drivers circling to find a space. This results in time wastage, increase in greenhouse gas emissions, loss of revenues to the cities, no parking, long waiting parking queues etc. Hence the solution is to gather the data from sensors in the cities and to develop Smart parking in cities to overcome the above issues. Moreover, it also helps cities for long term planning of resources [3].

The paper is organized further as follows. The 2nd section reviews the literature resource and gives chosen case studies, while the section 3 states the problem and describes the overall goal and needs of investigation. Thus, the section 4 provides the most important of our work – the analysis and set of experiments achieved. Main orientations at each step encompass both traffic and Smart City management, respectively.

2. LITERATURE REVIEW AND CASE STUDIES

2.1 Literature review

Li Hao et al [1] stated the growth and enhancement of Internet of things, Information and communication technology (ICT) and cloud computing as the keys factors in the development of Smart Cities. The major goal is to achieve the high-performance city facilities and operations, high standard of living, healthy environment and security. It focuses on characteristics, development, construction and trends of Smart City. Appling next generation of ICT, it deals with various aspects of the city problems like water, electricity, transport etc.

Irfan Ahmed Halepoto et al. [2] presented the continuous increase in population as the major challenge to the urban cities to maintain their resources with the expansion in the demand. Thus, they highlighted the need to develop intelligent cities to manage and preserve their resources. This paper also discusses the concept of SWOT (Strengths-Weakness-Opportunities-Threats) analysis, defined as a tool to measure the difference between existing and required resources in the Smart City. Including technologies like integration of cloud computing, next generation ICT, internet of things, it is demonstrated that Smart City utilize various algorithms of artificial intelligence and machine learning to function in the efficient way. Moreover, the presence of the dynamic technology for data collection, advance signal processing and data analytics tool, is mentioned as the major strength. In the same way, the existence of tools like GPS and remote sensing platform gives all the details for the city geographic locations and their respective details; which data gives an opportunity to the Smart City to perform development and implementation on wide scale, space and time service platform. The paper concludes that SWOT analysis is a strategy tool in transforming a city to a Smart City.

Paper [3] described how ICT is becoming widespread to support the Smart Cities to raise the living standard of the inhabitants. It also discusses the big data storage and data analytics which indeed generate vital knowledge for making predictions and decision making for the future. Thus, it is acknowledged that with the continuous increase in urban population, ICT is used to impart useful knowledge that is used for social and economic growth for better city management, playing so a vital role in improving Smart City governance by enhancing its facilities and conserving its resources.

The paper [4] is a Smart Cities guide from GSMA, which describes how Mobile networks and services can help municipalities to improve and manage transport through a city, whether it be by car, bus or on foot. Considering the change of Mobility as passengers and travelers are given more options and more flexibility; or the change of city management, as citizens become more demanding and employ more diverse modes of transport; transports options in Smart City are predicted to become much more efficient and relevant using connected sensors, big data, apps and mobile payments. Then, it advises on how to meet the transportation challenges of the future.

2.2 Case studies

The first case study belongs to Smart City in Shanghai. This is the NB-IoT connectivity which enables drivers to search for and book parking spaces, navigate their way to an allocated parking space, pay for the parking directly using a mobile handset, and manage their parking needs with mobile phone apps. This solution is designed to help tourists find parking spaces and relieve traffic pressure in the surrounding streets. The NB-IoT connectivity enables drivers to search for and book parking spaces, navigate their way to an allocated parking space, pay for the parking directly using a mobile handset, and manage their parking needs with mobile phone apps. China Unicom is already considering further cooperation with industry players to develop NB-IoT applications in new areas such as smart metering, crowd management and environmental monitoring.

The second case study is the Far EasTone Cellular Vehicle Probe (CVP) system in Tainan, which tracks the routes and origin-destination pairs of anonymous users, which can be used to monitor the level of service of roadway systems and identify traffic demand patterns for heavily congested bottlenecks. This information is then used to dynamically fine-tune traffic signal timings or to identify how best to change road layouts. The CVP system stores historical data for three years and can be used to retrospectively evaluate implemented transportation policies.

3. PROBLEM STATEMENT

3.1 Smart Parking

Parking-related issues can be high on smart cities' agendas. Whether it be reducing parking spaces to encourage citizens to use other means of transit, maximizing the use of existing parking spots, or ensuring that all parking revenue is accrued, smart parking solutions have a role to play. Integrated systems can highlight available parking spaces to drivers and route them to the parking spot. The data from connected smart parking sensors can also be used to enable dynamic pricing.

If a parking lot is empty, pricing can be adjusted to encourage more people to park there. This can also be linked to other data from connected sensors such as those measuring air quality – if air quality is poor, parking prices can be raised to discourage people from driving into the city Centre.

Introducing smart parking solutions help urban planners maximize their use of available space and ensure that parking does not dominate a city. Careful use of the technology enables both citizen engagement and timely enforcement, ensuring that parking space revenue and utilization is maximized. Smart parking solutions can also curb congestion by reducing the number of drivers on city roads who are simply looking for parking spaces.

Smart parking solutions can help venue owners, as well as city transportation departments. Supermarkets, stadiums and other venues can make use of smart parking solutions to ensure that their car parking is not abused.

Smart parking solutions can take one of several forms. Most common is the use of connected sensors embedded into the pavement to register if a vehicle is present in the parking space above them. Other sensor technologies can be used; as can cameras have connected to license plate reader platforms. [4]

Other connected sensors can also help build up a picture of parking availability – sensors placed on the entrance and exit to a car park, as well as video analysis of empty parking spaces can give a good idea of how many parking spaces are free and can also be integrated into a city's dashboard to provide real-time reporting. This is the approach we will use in this paper.

The aim then is to analyze use cases for Smart City parking which can help the inhabitants in the city to park their vehicle without any frustration and delay and reduce greenhouse effect.
Some of the use cases developed for Smart parking are:

- Reduce time to find parking space
- Regulate parking and traffic in the city
- Identify area to build new parking
- Dynamic parking pricing
- Parking space availability prediction

With average driver spending 6 to 14 minutes, in searching for parking, parking is a challenging issue in cities; hence 30 percent of all traffic congestion is caused by drivers circling to find a space [7]. This results in time wastage, increase in greenhouse gas emissions, loss of revenues to the cities, no parking, long waiting parking queues etc. Thus, the solution is to gather the data from sensors in the cities and to develop Smart parking in cities to overcome the above issues.

Smart parking imparts visibility of parking by giving details like usage and vacancy of parking garage. It improves occupancy of the parking garage and helps the flow of traffic in the smooth manner. Moreover, it also helps cities for long term planning of resources.

3.2 Traffic management

Smart city solutions can help address city congestion: Connected sensors and mobile data can be used to both monitor congestion and to also direct traffic in dynamic ways to prevent congestion building up or getting worse.

Connected sensors for traffic management can be placed in a wide variety of locations and can collect a wide variety of data. In-vehicle sensors can monitor their speed and location. Sensors can also be placed on the street, with license plate readers and CCTV cameras offering a good view of congestion in real-time. Sensors can also be integrated into the road surface itself, counting the number and type of vehicles passing overhead in any given timeframe.

A combination of these sensor types gives a very accurate picture of traffic levels, and allows also a wider view across the city, rather than at just one specific location.

Today's traffic management technologies typically only measure traffic at specific locations, and do not have the intelligence to see what is causing the congestion further up the road. In the past, these systems have had to be supplemented by citizens calling in to report issues that need resolving. In many cities, existing sensors are at the end of their lifespan, and can be replaced with connected sensors with minimal disruption. If using sensors connected by a mobile operator, the real-time data across a wide area of the city can be monitored and appropriate actions taken. By installing further connected sensors in street furniture, such as traffic lights and road-sign displays, vehicles can be directed down different routes or congestion cleared by turning traffic lights green.

A fully dynamic system can monitor traffic and take appropriate actions, such as automatically turning lights green. By bringing all the available data from sensors, cameras and city employees into a single point, quick decisions can be made and acted upon. Thus, connected systems enable three approaches to traffic management:

1. Preparation for known changes, issues and restrictions, such as roadworks or road closures, whereby the city is able to inform people travelling within the city of changes, while using connected traffic lights and signs to re-route vehicles away from the affected area;
2. Dynamic changes in response to incidents, such as accidents or weather problems;
3. Urban planning. By looking at accurate traffic trends, urban planners can permanently alter the roads infrastructure to reduce traffic.

Our focus in this paper demonstrates the latest 2 concerns. The number of vehicles on the roads are increasing rapidly. It will lead to enormous rise in traffic congestion, accidents and pollution or several situations listed above. It raises one of the vital questions of how to handle more vehicles with less traffic congestion on highways and in cities in the next few years. In addition to traffic jam it also increases pollutant level causing health problems in the city.

Hence there is a need to manage traffic flow in the city collecting and analyzing the data based on different patterns such as traffic speed, density, speed limits, timestamp, location etc. Based on this information city can be better planned having balance mobility, safety and reducing environmental issues like fuel emissions. The goal of this paper is then, also to demonstrate through use cases how traffic management in the city helps us to route the traffic in the city in more efficient way. Some of the use cases that are implemented for Smart City traffic management are:

- How to identify less used roads during rush hours?
- How to identify points where traffic jam appears?
- How to predict traffic jam based on traffic patterns?
- How to reduce time wasted in slow traffic?
- Determine where and when to build new roads, bridges

The data is analyzed to find the data insights for the similar use cases. On this basis, traffic lights can be controlled on the roads which were busier to avoid congestion and traffic jams. New bridges and alternate routes can be constructed to handle high volume of traffic which will further reduce accidents.

4. EXPERIMENT ANALYSIS

In this section, we provide few results of subsequent use cases implemented. Data are collected from different sources according to the use case requirement. These results include both preliminary analysis and more deepened analysis such as prediction and Smart Parking or Traffic monitoring.

4.1 Experimental environment

The setup environment is fully based on opensource software and it is complying with the Analytics Implementation Architecture. The environment used in an Hadoop Cluster environment with fully features; there is a gap of options such as Cloudera solution, Hortonworks solution or a native MapReduce-Yarn/HDFS solution built from scratch.

The most important tools from the features of Hadoop solution we needed for our experiment are Hive for retrieving data from the Hadoop data file system (HDFS). We also needed Python which has positioned itself as one of the best Programming languages having the most requested libraries in data wangling or analysis. Python was installed on Linux machine more suitable for Python modules and Packages integration.

Apache HTTP server was also needed to remotely execute webpages residing at server side from client. Its configuration file httpd.conf is reconfigured to execute python scripts, so that webpages can call python scripts and fetch results to display on GUI.

4.2 Design and Analysis

4.2.1 Traffic area and objects under experiment

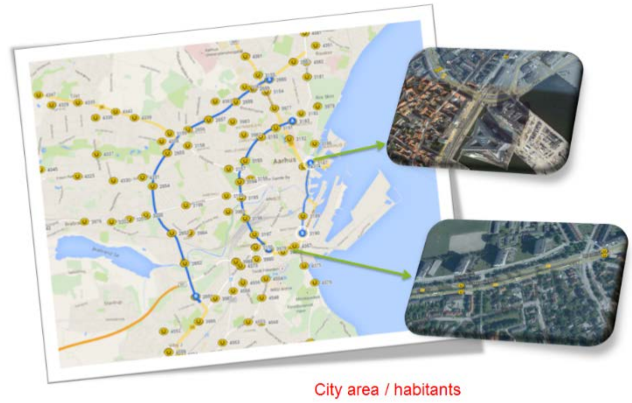

Figure 1. Aarhus Traffic Observation Points.

Data is collected for vehicle traffic provided by City of Aarhus in Denmark, which have 449 traffic reporting pairs, build out of 136 single observations as shown on the figure 1 below. The goal of this project is to develop use cases for traffic management in the city that help us to route the traffic in the city in more efficient way.

The figure shows the 136 traffic points in the city of Aarhus on the google map. Preliminary analysis is done by observing the traffic pattern in between the two traffic points.

4.2.2 Preliminary Analysis

Figure 3 shows the traffic details between points 3193 and 3194. It can be noticed that rush hours seem to start around 3 am, and traffic Jam appears around 6 am on February 27th, 2014. Here the preliminary analysis is done to know the traffic pattern between two points throughout the day.

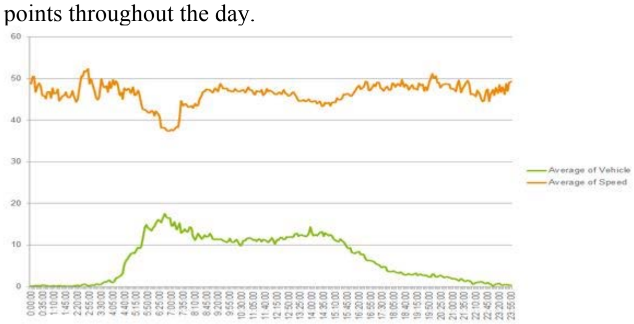

Figure 2. Traffic Details between Points 3193 and 3194.

Obviously, less we have vehicules on a road, more the speed increases; similarly, more the number of vehicles on the road grows, deeply decreases the speed of their movement, conducting to a possible traffic congestion.

The next analysis encompases available places in the garages under study and abnormal values which can be noticed. Figure 3 shows the preliminary analysis done for the eight parking locations for the city of Aarhus in Denmark. It shows all the eight parking garages with their maximum capacity, maximum number of vehicles and average count of the vehicles in the garage.

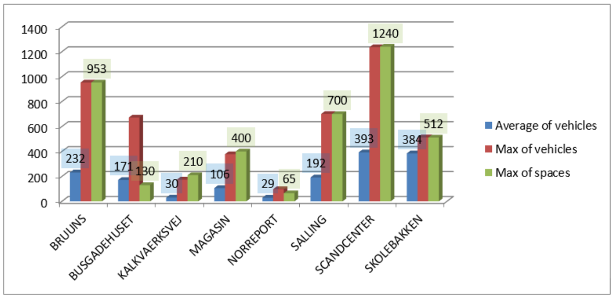

Figure 3. Preliminary Analysis for Parking Garages

From the above analysis it can be observed that:

- In average, 60 % parking places are unoccupied
- Parking reach full capacity during the observation period
- Anomalies in the datasets are human observable

It can be observed that collected data have errors due to sensor problem. It is noticed that Parking garage BUSGADEHUSET is showing more number of vehicles then its maximum capacity.

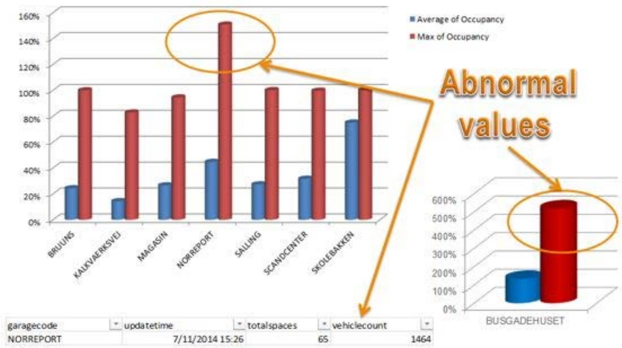

Figure 4. Occupancy Rate for Parking Garages

Figure 4 gives us the details of the error in the collected dataset. It is showing the maximum capacity of the eight-parking garage and their maximum occupancy. Hence it is vital to identify relevant data, cleaning, filtering, joining, grouping and transforming the data to find its data insights.

4.2.3 Deeply Analysis

Getting deeply in analyzing these data through integration and correlation with multiple events occurring in the city, it is noticed that, the occurrence of the event has direct correlation with the nearby parking location.

In general, Python scripts fetch results from database, Hive in our concern, for the use cases and create graphs using python 2D plotting library. Webpages are created using HTML and JavaScript to display the dynamic change in data with respect to timestamp to get a real-time view of the application. JSON (JavaScript Object Notation) is used as the lightweight data-interchange format between HTML and Python. JSON data files are also created from python scripts as the intermediary results. These files are created to improve performance of the working application. Depending upon the data size, querying big data might take long to give the results back. So, these JSON files are used for better performance.

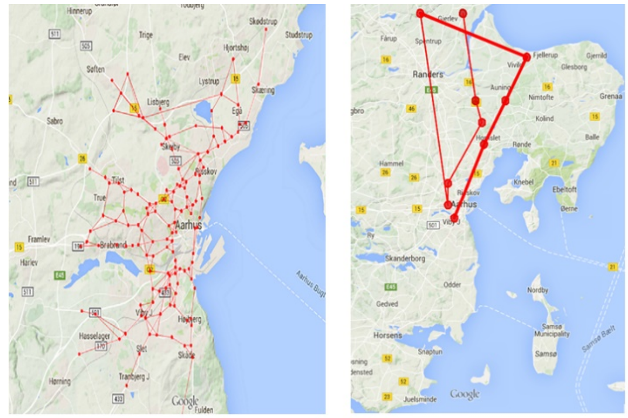

Figure 5. Flow of Traffic in the Aarhus City

Our deeply analysis focuses on prediction. Figure 5 (B) shows the flow of traffic for ten traffic points. Flow of traffic in between two points depends upon the vehicle count. Number of vehicles travelling is shown by the thickness of the path. Real time application is implemented by taking the number of vehicle count from the database and displayed on the webpage. Change in vehicle count is dynamic on the webpage with respect to the timestamp.

Figure 6. Parking garage correlation with events

Figure 6 shows the second Smart City parking use case. This use case shows the parking garage correlation with the events occurring in the city. The data for the two event locations and three parking garages are analyzed. It can be noticed that with the occurrences of the event the space available in the parking goes down with the increase in the parking price. When the parking garage has less than 75% of used parking occupancy it is shown as green icon. When it's used occupancy reaches 75% or more then it is shown as orange icon and when there is no parking space available in the garage then it is shown as red icon. Webpage is updated dynamically with time. The details of the parking garage and event location in the infowindow such as parking garage name, time, space available and current price of the parking, are shown.

5. ACKNOWLEDGMENTS

Our thanks to Ericsson Canada for giving me the opportunity to use their development infrastructure during early stage of this research.

6. REFERENCES

[1] Li Hao et al., "The application and implementation research of smart city in China", "System Science and Engineering (ICSSE)", Dalian, Liaoning, 2012, © IEEE, doi:10.1109/ICSSE.2012.6257192

[2] Irfan Ahmed Halepoto et al., "Multi-criteria assessment of smart city transformation based on SWOT analysis", "Information Technology: Towards New Smart World (NSITNSW)", Riyadh, Saudi Arabia, 2015, © IEEE, doi:10.1109/NSITNSW.2015.7176412.

[3] Zaheer Khan et al., "Cloud Based Big Data Analytics for Smart Future Cities", "Utility and Cloud Computing (UCC)", Dresden, 2013, © IEEE, doi:10.1109/UCC.2013.77.

[4] GSMA, Connecting Living. GSMA Smart Cities Guide: traffic Management. Why mobile operators are key partners for cities seeking to deploy sustainable traffic management solutions, 2016, © GSMA.

[5] Vangipuram Radhakrishna, Shadi A. Aljawarneh, Puligadda Veereswara Kumar, and Kim-Kwang Raymond Choo. 2018. A novel fuzzy gaussian-based dissimilarity measure for discovering similarity temporal association patterns. Soft Comput. 22, 6 (March 2018), 1903-1919. DOI: https://doi.org/10.1007/s00500-016-2445-y

[6] Shadi A. Aljawarneh, Ali Alawneh, and Reem Jaradat. 2017. Cloud security engineering. Future Gener. Comput. Syst. 74, C (September 2017), 385-392. DOI: https://doi.org/10.1016/j.future.2016.10.005

[7] Shadi A. Aljawarneh, Radhakrishna Vangipuram, Veereswara Kumar Puligadda, and Janaki Vinjamuri. 2017. G-SPAMINE. Future Gener. Comput. Syst. 74, C (September 2017), 430-443. DOI: https://doi.org/10.1016/j.future.2017.01.013

Authors' background

Your Name	Title*	Research Field	Personal website
Mama Nsangou M.	**Affiliate Assistant professor**	**Computer Science and Software Enginering**	**www.linkedin.com/in/nsangou-mouchili-91703133**
Shadi Aljawarneh	**Full professor**	**Software Engineering**	**www.linkedin.com/in/shadi-aljawarneh-7a799419**