Conference on Small and Medium Smart Cities
Alcoy, 14-15 February 2018
(pp. 1–5)

How Data Science helps to build Smart Cities: València as a use case

Lidia Contreras-Ochando[1], Cristina I. Font-Julián[2], David Nieves[1], Fernando Martínez-Plumed[1]

[1] *Departamento de Sistemas Informáticos y Computación, Universitat Politècnica de València, Camí de Vera s/n, 46022, València. E-mail:* {liconoc,daniecor,fmartinez}@dsic.upv.es.
[2] *Departamento de Comunicación Audiovisual, Documentación e Historia del Arte, Universitat Politècnica de València, Camí de Vera s/n, 46022, València. E-mail:* crifonju@upv.es.

Keywords: Data Science, Machine Learning, Open Data, Smart Cities

Resumen

The high degree of datification and connectivity embedded in a Smart City demands tools and mechanisms for data manipulation, knowledge extraction and representation that facilitate the extraction of meaningful insights. Clearly, Data Science can make enormous contributions to the development of Smart Cities, especially when it comes to gather and process information, combined with the capabilities of machine learning. In this regard, this paper discusses the use of Data Science methodologies and machine learning techniques to Smart City management aspects such as infrastructures, public safety and health, citizens' empowerment, transportation, etc. and presents a number of practical cases in the context of Smart Cities in València, Spain.

1. Introduction

Smart City projects, which main objective is to enhance the well-being of citizens, have gained momentum over the last decade. Thousands of municipalities worldwide have launched a wide variety of digital transformation initiatives aiming at changing citizens lives at many levels such as less pollution, garbage, parking problems and more.

València (Spain) is a clear example of a Smart City, with many technological and open data initiatives. It has been a pioneer city in Spain which has centralised all of its municipal information through a Smart City technological solution: *València Smart City platform*. This platform provides several portals, applications and open data repositories which allow citizens, policymakers and private companies to get insight, in real time,

This work has been partially supported by the EU (FEDER) and Spanish MINECO grant TIN2015-69175-C4-1-R, by Generalitat Valenciana PROMETEOII/2015/013, and by Conselleria de Transparència, Responsabilitat Social, Participació i Cooperació

about the essential information of the city and, therefore, to have the capacity for taking action and anticipating situations that affect the daily life of citizens. The information provided ranges from public transport, emergency and security services, public facilities, environment, cleansing, waste collection, lighting, crane service, gardening or meteorology.

This high degree of datification and connectivity embedded in a Smart City such as València, demand tools and mechanisms for data manipulation, knowledge extraction and representation that facilitate the extraction of meaningful insight. Data Science can make enormous contributions to the development of Smart Cities, especially when it comes to gather and process information. Furthermore, Machine Learning (ML) and Artificial Intelligence (AI) algorithms can be applied to open data to produce and extract meaningful knowledge for Smart City applications and, therefore, to lever the development of personalised services. On this matter, not only have we focus on the data acquisition, transformation and cleansing processes already from open data repositories, but we have also studied and adapted well-known ML and AI algorithms which have proven to be instrumental when it comes to make estimations regarding the use of different means of transport (bikes, traffic, ...), *(b)* improve infrastructures (cycling networks, ...), *(c)* predict and extrapolate pollution and pollen levels, and *(d)* detect patterns of fraud. This overall process has resulted in the development of several Smart City projects and approaches that will be briefly described in the following sections.

This short paper is structured as follows: Section 2 briefly describes the methodology followed to develop our Smart City projects. Section 3 outlines the different projects, their objectives and goals. Finally, Section 4, concludes the paper.

2. Data Science for Smart Cities projects

Since the amount of data generated is growing exponentially as the number of this type of smart devices and open data platforms increase, new knowledge extraction methodologies and techniques have to be used. In this regard, Data Science provide us a suitable framework within which we can confront the challenges of developing intelligent projects and applications. In this context, data extraction, selection and transformation processes as well as the use of analytical and visualisation tools are of the utmost importance to succeed in our efforts to obtain valuable insight and knowledge from these sources. This data pre-processing phase is, thus, one of the first and critical steps to data mining and data analysis which outputs are directly inputted to our intelligent Smart City projects as predictive or descriptive models. All the projects we have carried out involve a substantial amount of effort w.r.t. data integration, cleansing, parsing, correcting, standarising, matching, reduction and selection.

Different data such as traffic intensity, pollution and pollen levels, bike sharing services demand, meteorological conditions, census, cartography, public and government spendings (provided by the city of València[1] and others organisms[2] [3]) has been analysed for each project (see Figure 1), having in mind not only to monitor, but to manage and improve the normal functioning of several city services. In this sense, the use of ML and AI techniques bring new opportunities to develop applications to estimate and extract patterns and trends (from the pre-processed data) in terms of mobility and transport of citizens, in-

[1]Valencia open data platform: `http://gobiernoabierto.valencia.es/es/data/`

[2]GVA data platform: `http://www.agroambient.gva.es/web/calidad-ambiental/datos-on-line`

[3]Spanish Observatory of R&D (ICONO): `https://icono.fecyt.es/pitec`

Figura 1: Projects developed (described in Section 3), grouped by their ultimate goal, their data sources, whether the perform a descriptive tasks (use of data aggregation and data mining to provide insight into the past) or a predictive (use of forecasts techniques to understand the future) tasks, and techniques used.

frastructure use, pollutants and pollen dispersion, or government expenditure. In general, different ML and AI techniques have been used for each project. Specifically, a subset of ML ensemble-based techniques such as Bagging, Random Forests and Boosting techniques has been selected, and particularly tuned, to improve learning methods based on decision trees (classification techniques) or linear models (regression techniques).

3. Projects

During the last years we have been working on a number of Data Science projects related to Smart Cities. The knowledge extracted from the aforementioned open data platforms has been used to develop specific applications aiming at improving multiple aspects or services in València. In the following, we briefly describe each application (objectives, data and, when applicable, prizes won) grouped into different application areas.

3.1. Optimizing infrastructures for cities

Infrastructure systems can be viewed as an opportunity to shift cities onto a more sustainable and efficient path by paying close attention to the resources that pass through them. In this sense we are working on several projects related to traffic congestion, bike sharing systems, design of bike lane networks and public services coverage along the city:

- **TrafficXplorer** (`http://www.dsic.upv.es/~flip/trafico/`): A web application to show and predict traffic levels. The application has three different sections: historical data (showing traffic intensity from one to twelve hours before current hour); real-time data (showing real-time traffic data); and prediction (showing the traffic intensity prediction from one to twelve hours after the current hour). 1^{st} prize in the I *OpenDatathon ETSINF-UPV 2016: BigML Award for Best predictive project.*
- **BikeXplorer** (`http://www.dsic.upv.es/~flip/bikeXplorer/`): An application to predict the demand of Valenbisi[4] stations. We use data about the historical demand of the different Valenbisi stations in the city. The application allows the user to obtain an estimation of the future demand in each Valenbisi station as well as the best route to reach a target station. In addition, the system is able to show the potential use of a new station in any point of the map, depending on the type of neighbourhood and census data. $1^{st}/3^{rd}$ local/national prize in the *IV Hackathon Telefónica HackForGood 2016.*
- **BikeFlow**[5] (`http://safe-tools.dsic.upv.es/bikeflow/`): Improvement of BikeXplorer. We enriched the Valenbisi demand with meteorological data characterizing each

[4]Valenbisi is a bike sharing system from the city of València (`http://www.valenbisi.es/`)

station independently. The application allows the user to find historical and real-time information and the estimation of future demand in each station. We also use extrapolation techniques to estimate the demand that a currently non-existent station at any point in the city would have. 1^{st} prize in the II *OpenDatathon ETSINF-UPV 2017: Valencia City council Award*.

- **ViceBisi** (`http://safe-tools.dsic.upv.es/vicebisi/`): This project provides an intuitive way to visualise and compare varying details regarding the bike services through the neighbourhoods of the city. Crossing different data sources, we generate a number of indicators (bike lane covarege, bike anchors per inhabitant, bike station distances, etc.) to reflect the coverage of these services for each neighbourhood. 1^{st} prize in the I *OpenDatathon ETSINF-UPV 2016: Valencia City council Award*.
- **MoreBikeLines** [4] (`http://users.dsic.upv.es/~flip/RutasBici/`): An approach to design and calculate bike lane networks based on the use of open data about the historical use of a urban bike rental services. Concretely, we model this task as a network design problem (NDP) and we study four different optimisation strategies to solve it. The proposed method can be easily used to improve or extend bike lane networks based on historic bike use data in other cities. 1^{st} prize in the I *OpenDatathon ETSINF-UPV 2016: BigML Award for Best predictive project*.

3.2. Improving public safety and health

In general, overpopulated cities around the world have common problems in areas such as health care and public safety. Examples of major problems in cities are the levels of pollution and pollen allergies. In this respect, we are working on a couple of projects addressing these two issues in the particular case of València:

- **airVLC** [1] (`http://safe-tools.dsic.upv.es/airvlc/`): The data collected related with air pollution is published with a three-hour delay, making difficult to prevent high levels. airVLC is an application that predicts the levels of pollution by employing traffic and meteorological data in real-time, providing different maps to show the level of pollution in each point of the city by using wind-based spatial interpolation techniques. 3^{rd} local prize in the *III Hackathon Telefónica HackForGood 2015*.
- **myAllergenApp** (`http://safe-tools.dsic.upv.es/myAllergenApp/`) An ongoing work to forecast pollen pikes and its spreading. It is a significant challenge due to the low number of sampling stations along the city and the delay in time to get the latest data. Given the above constraints, several data analysis and ML models are being developed to estimate pollen levels across the city, as well as to forecast allergens peaks using data ranging from tree census, meteorology, pollen or seed dispersal data.

3.3. Citizens empowerment

In the processes of citizen empowerment, individuals and communities are mobilized in the political and social spheres, to enhance their living conditions and, thus, construct a more democratic, equitable and just society. In this context we are working in the development of systems to improve expenditure control and fraud screening:

- **SALER** [3] (`https://safe-tools.dsic.upv.es/saler/`): Technically known as the *Security Administrator Tool to Analyze Networks* (SATAN), this ongoing official project aims at detecting bad practices and fraud in public administration. The system analyses data from the databases about procurement records, imprest funds, grants and staff info from the *Generalitat Valenciana* with the aim of extracting patterns of fraud as well

as correlations and associations between data and descriptive attributes. Hervé Falciani and the *Fundación Baltasar Garzón* have participated in its development.

- **Transparency Science** [2] (`http://dataupv.webs.upv.es/transparencyscience`): The aim of this project is to provide reliable information about public investment in science. Data related to the Spanish state budget is collected from several public open sources. Once processed, the knowledge extracted is shown in such a way to facilitate maximum understanding of citizens, as well as their participation by means of voting systems. Finalist on the international *iConference's Social Media Expo 2015*.

3.4. How People Use Cities

Being able to analyse the citizens' movements within a city, as well as how or the frequency of their move, can help to improve not only the services or infrastructures needed to fulfill the citizens needs, but to improve the city as a whole. A couple of projects have been developed in this sense:

- **Traffic@Halloween** (`http://liconoc.webs.upv.es/halloween/`): A study about the traffic flow of València in the Halloween's evening and night (2015). It provides information about different traffic parameters (movement, vehicles per hour) in the streets of the city at different periods of time. The knowledge extracted and visualisations involve the most visited zones for dinner and party during the Halloween night.
- **Traffic@Tarongers** (`http://liconoc.webs.upv.es/tarongers/`): This study was carried out in concert with the UPV as an allegation to the Valencia urban plan presented by the city council in 2015[5] in order to show that the *Tarongers avenue* (Valencia) is not one of the most used avenues in València. We used historical traffic data to analyse and show (visually) which areas have huge traffic and congestion problems.

4. Conclusions

In this paper we show that Data Science methodologies, together with ML and AI techniques, provide us with the appropriate tools to address the challenges of developing intelligent projects and applications for Smart Cities that go beyond mere data analysis and visualisation projects. To illustrate this, we have introduced the reader to a number of projects (including pointers to further information about them when possible) developed in the context of the *València Smart City platform*.

References

[1] L. Contreras-Ochando and C. Ferri. Wind-sensitive interpolation of urban air pollution forecasts. In *International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA*, pages 313–323, 2016.

[2] L. Contreras-Ochando, C. I. Font-Julián, P. Morillo, and D. Vallejo. Transparencyscience: Return on research investment, where do the funds go? *iConference 2015 Proceedings*, 2015.

[3] EFE. Un sistema de alertas vigilará irregularidades para evitar casos de corrupción. *EFE Agencia*, 3 Nov 2017.

[4] F. Martínez-Plumed, C. Ferri, and L. Contreras-Ochando. Cycling network projects: A decision-making aid approach. In *Proc. of the 1st Data Science for Social Good co-located with European Conference on Machine Learning, SoGood@ECML-PKDD 2016, Riva del Garda, Italy, September 19*, 2016.

[5] J. Sánchez-Martínez and F. Martínez-Plumed. Bikeflow: aplicación para la predicción de la demanda de bicicletas en la ciudad de valencia. *II OpenDatathon ETSINF-UPV: TechReport*, 2017.

[5] `http://www.valencia.es/planos_urbanismo/NORMATIVA/PGOU/Normas_web/normas.htm`