How Machine Learning is Changing e-Government

Charalampos Alexopoulos  
University of the Aegean  
81100, Mytilene  
Greece  
alexop@aegean.gr

Zoi Lachana  
University of the Aegean  
81100, Mytilene  
Greece  
zoi@aegean.gr

Aggeliki Androutsopoulou  
University of the Aegean  
81100, Mytilene  
Greece  
ag.andr@aegean.gr

Vasiliki Diamantopoulou  
University of the Aegean  
81100, Mytilene  
Greece  
vdiament@aegean.gr

Yannis Charalabidis  
University of the Aegean  
81100, Mytilene  
Greece  
yannis@aegean.gr

Michalis Avgerinos Loutsaris  
University of the Aegean  
81100, Mytilene  
Greece  
mloutsaris@aegean.gr

ABSTRACT

Big Data is, clearly, an integral part of modern information societies. A vast amount of data is daily produced and it is estimated that, for the years to come, this number will grow dramatically. In an effort to transform the hidden information in this ocean of data into a useful one, the use of advanced technologies, such as Machine Learning, is deemed appropriate. Machine Learning is a technology that can handle Big Data classification for statistical or even more complex purposes, such as decision making. This fits perfectly with the scope of the new generation of government, Government 3.0, which explores all the new opportunities to tackle any challenge faced by contemporary societies, by utilizing new technologies for data-driven decision making. Boosted by the opportunities, Machine Learning can facilitate more and more governments participate in the development of such applications in different governmental domains. But is the Machine Learning only beneficial for public sectors? Although there is a huge number of researches in the literature related to Machine Learning applications, there is lack of a comprehensive study focusing on the usage of this technology within governmental applications. The current paper moves towards this research question, by conducting a comprehensive analysis of the use of Machine Learning by governments. Through the analysis, quite interesting findings have been identified, containing both benefits and barriers from the public sectors’ perspective, pinpointing a wide adoption of Machine Learning approaches in the public sector.

CCS CONCEPTS

- Applied computing → Computers in other domains → Computing in government → E-government

KEYWORDS

Machine learning, government 3.0, government services, big data, artificial intelligence

ACM Reference format:


1. INTRODUCTION

During the recent years, the explosion of digital data has become a reality and the ability, for public and private sector on collecting and processing them has increased dramatically. This led to a new era of digital data and a new trend, the ”Big Data” appeared. In fact, during 2000 the data flood only from Google’s indexed web pages increased 1 million times and until 2008 has reached and soon after exceeded 1 trillion [23] compared to the 1998’s data flood. The vast amount of data consists a consequence of the acceptance, especially, of new emerging technologies, including their associated devices (such as Inter-
of Things (IoT) applications which facilitate the generation of data through applications, sensors, etc.) [24] and social net- working applications such as Twitter, Facebook e.t.c. (allowing users to create content and amplify also the huge web volume) [25]. But collecting data is not enough. When someone refers to “Big Data” we can easily understand that the capabilities of processing, analyzing, storing and even understanding these datasets is impossible without letting the machines help us. It is argued that the key on processing Big Data is the Artificial Technology (AI). AI consists the way of transforming computers into intelligent machines [4] through the abilities of learning and perception [25].

At the same time, and with the broader evolution of electronic government, whose main scope [1] is to integrate the new, disruptive ICT developments with the established ICTs for policy making and evidence-based decision, new trends have emerged in the public sector. Big Data, Blockchain Technologies, AI and particularly ML Algorithms are some of the technologies used for modernizing the previous services providing by governments [1, 2]. For many years, there has been an extensively use of AI techniques for enhancing decision making quality and for solving problems in different industries. The utilization of a variety of machine intelligence types, such as robotics, natural language comprehension, ML and neural net- works, can contribute to this purpose [63]. Generally, ML can be defined as the field of study that allows the machines to learn with no use of explicit programming [3]. As Abbod et al. mentioned [27] “Learning can be used to train a machine, so that it optimizes its rule base in a model and then new parameters maybe tested in that model”.

Previous research mapping has indicated that, various ML techniques are being used, in the public sector [3]. The usability of them can be met also into different domains in order for an automated data analysis to be achieved due to the sheer of volume of the accumulated data [1, 2]. In fact, tools suitable to analyze and explore big data include ML techniques, and in particular surrogate models [77]. Chui T. K. et al. [3] and Pereira et al. [1] mentioned that a variety of complex problems addressed in contemporary societies can be tackled through the use of ML techniques (such as classification, decision making and face detection). However, a series of challenges exist in the utilization of ML in the e-Government field. Therefore, a further analysis of ML is required, including ML’s capabilities and benefits, in order to determine all key factors of the adoption of this technology in the public sector. This research contributes towards this purpose, as a review in the existing literature for all key characteristics of ML is implemented, including also the current landscape of ML applications in the public sector. The outcomes of this analysis, including specific cases of ML projects in the public sector, should provide a clear understanding and estimation of the added value of adopting ML in government. Hence, this study aims at the identification of the benefits and barriers that the public sector may encounter by the adoption of ML, with the view of shaping all the directions for future research in the field of Government 3.0.

The rest of this study is structured as follows: In section 2 background information is provided, necessary for landscaping the field we examine. Section 3 presents our methodology which underlies our research. In section 4 a literature review is presented while sections 4 and 5 present our research’s results. Finally, in section 6 suggestions and further research steps are presented.

2. BACKGROUND

The term “Big Data” was first appeared by John Mashey [26] in 1998 and, during the same year, the first book [22] mentioned that Big Data and mining were relevant, was, also, revealed [23]. Two years later, we can find the first paper with the term Big Data written in the title [26]. However, as Gandomi and Haider [45] mentioned, the term Big Data was intensely spread after 2011 with a peak on 2013 and until 2015 there were many different definitions focusing either on what Big Data does or what it is. Big Data is, also, characterized by 6V’s; i.e. Volume, Variety, Velocity [49, 48, 50, 51, 23], Variability, Value and Veracity [48, 45, 23]. We can realize that Big Data is a new term and refers to the datasets which are difficult, or even impossible, to manage with the use of traditional and common ways or even simple data mining tools. Specifically, except Big Data Analytics [46, 43, 47, 48], Big Data Mining [23, 42] can, also, be used for creating value of the data by extracting useful information from these large datasets or streams of data. E-Government could achieve its next generation towards data-driven and evidence-based decisions and policy-making [1], by trans- forming Big Data into useful information through advanced data processing techniques. Considering also the huge amount of big, open and linked governmental data, we can conclude that advanced techniques can be considered as a requisition for public sector to complete their goals.

Big Data Computing is a powerful paradigm of enabling scientists on tackling challenges in many different disciplines and is, also, a market of many billion dollars [30]. ML consists an important technology for processing Big Data. ‘Big Data Machine Learning” has been used widely in the fields of both pub- lic and private sector. ML has flourished in the ‘90s [56] and was first used in the field of Statistical Science, while there were made attempts to learn a computer how to play games. In this period of time, ML algorithms were designed for a variety of purposes, such as speech recognition, but also for providing data-driven answers to vexing questions [10]. According to Chui K. T. et al. [3], there are four types of ML: Un-supervised learning (unlabeled training data), which commonly applied algorithms for supervised learning are Decision Trees, Support Vector Machines (SVM), Naïve Bayes, Neural Networks and Maximum Entropy (ME), Supervised Learning (labeled training data in type), Semi – supervised Learning (few labeled and many unlabeled data in type) and Reinforcement Learning – a totally different type and its main goal it to learn how to control data [54, 3]. ML’ s field is organized around three primary re- search objectives, Task – Oriented Studies, Cognitive Simulation and Theoretical Analysis [52]. Literature provides a huge amount of ML models and few examples among all are “classification and regression trees”, “neural network” (Multilayer Perceptor),"Bayesian Neural
network” , “support vector Regression, K- nearest neighbor model “(KNN) and “Gaussian Processes”.

There are many techniques in literature, which can be used for the development of an ML model. Petter Jeffcock [53] captures four of them:

- **Regression** - These algorithms are being used for numeric predictions
- **Classification** - Enables the membership detection among a known class
- **Clustering** - Logical Grouping of mass of data
- **Anomaly Detection** - Identification of rare/ different items among a dataset considering the data majority

The continuous advance of ML is crucial in many fields such as cybersecurity and scientific discovery as well as in multiple business domains [29]. For example, the topic modeling and the collaborative filtering algorithms (ML algorithms) are often used for the improvement of users’ experience and for revenue increasing [31, 32, 33, 34]. ML is used for information extraction from a raw of data and it can be used for a variety of purposes (e.g. prediction, understanding) [44]. Matt Leonard [78] pinpoints that over the past two years (2016 -2018) governments have increasingly outline Machine Learning as a research priority for a better understanding of governments data and implementing more efficient government solutions. When it comes to government, ML algorithms can help in the identification of significant factors and yet not defined interrelations and as such can be used to decrease the complexity of societal phenomena that are related with policy problems. To this direction, predictive modelling is defined as the analysis of large data sets to make inferences or identify meaningful relationships that can be used to better predict future events [69].

ML techniques in predictive modeling are used for the analysis of both current and historical facts for predictions making either for future or unknown events. For example, ML algorithms, in combinations with clustering techniques, can contribute to the structure of the online information, and the use of sentiment analysis (SA) can help policymakers for decision making, by informing them about any changes in public opinion, based on current political discussion trends. Also, Natural Language Processing techniques on structured and unstructured texts, with the view of extracting from large corpora (difficult to read), can extract comprehensible, timely and direct insights for people’s opinions, emerging issues, trends, behavioral, events against policy topics [70].

3. METHODOLOGY

The methodological approach of this research study is presented in the current section. In particular, a comprehensive review of the current landscape of ML was conducted in order to understand whether ML is beneficial, and which are the barriers hindering its adoption by the public sector. An initial review of the literature guided us to enlighten all the key characteristics of ML, with particular focus on the e-government solutions which use ML techniques. Particularly, the initial study was based on the cautious examination of the DGRL (Digital Government Reference Library, previously named Electronic Government Reference Library (EGRL)), the Google Scholar, the IEEE Xplore, the Scopus and the Web of Science bibliographic data- bases for relevant publications, using the keywords: “machine learning government”, “machine learning benefits barriers”, “machine learning public sector”, “machine learning review”.

In the aforementioned bibliographic databases, forty-three (43) in total research papers were identified and guided by these, the diverse application domains of ML applications in the public sector were pinpointed.

Three research questions have been answered by this first step of the methodology:

1) What are the general benefits of the adoption of ML by governments?
2) What are the general barriers to the adoption of ML by governments?
3) What are the general uses of ML implementations by governments?

Since the barriers are dominating (numerically) the benefits and due to their field of nature (the research revealed that most of the barriers hide upon the data) a further research is required to explore whether these barriers prevent governments of implementing ML solutions or not. Next, further research on Horizon2020, Erasmus+ and on Connecting Europe Facility (CEF) research funding framework programmes was held for the identification of existing research projects in the field of e-government, utilizing ML. The involvement of at least one governmental organization in each project’s implementation was considered as the main parameter narrowing the results. Through desk research, an analysis of the identified projects made by its Technology Readiness Level (pilot, proof of concept, large-scale implementation), application areas (according to Horizon2020 2 and Government areas 3), governmental bodies involved in the implementation of the project, and the scope of each project is presented. In the following sections, the above steps’ results (e.g., applications re-view) are reported.

The second step of the methodology we followed consist of a detailed presentation of the ML projects is conducted including, also, the countries (Governments, see Table 3) involved in the implementation of each application which is the result of the second step of the methodology.

Finally, the last step of the methodology aggregates the results from the above steps are brought together.

4. RESULTS

4.1. Applications of machine learning in government

This sub-section presents an overview of the applications of ML in governmental sectors, while the subsequent sub-sections present a list of the benefits and barriers of the ML that have identified in the research papers under study. In, particular
Several works have been found in literature, exploring or proposing technologies, including ML techniques, for analyzing automatically the data due to their structure and their sheer volume \[3\] for improving the relations among citizens and governments \[28\].

Various ML models are, also, being used in order to classify data from social media platforms, such as Facebook and Twitter, into predefined categories \[6\]. As Sakai and Hirokawa mentioned \[35\], ML techniques can be used as support vector machine (SVM) \[77\] and word feature selection (SVM + FS) in order to analyze citizens’ reports, for instance danger detection signs reports even by using social media reports or even for real danger detections – reports \[37\]. The previous techniques could achieve higher performance in detecting danger signs compared to humans considering, also, that judgment from humans has a low rate of agreement \[36\]. Also, a promising usage of ML on the same field can be detected for political purposes \[6\] or for evaluation of politicians’ truthfulness \[39\].

Tourism industry through forecasting visitor arrivals (demand indicator) can be served as a tourism reference for the public sector \[41\]. Using ML techniques, including more detailed data (such as road/air transport, accommodation, and art), can be beneficial for economic development \[41\].

In New Zealand, ML models are being used in livestock industries for the predictions of livestock estimation with biosecurity, broad applications in disease risk modelling policy and planning \[40\], while Wilbanks, J.T. & Topol, E.J. \[14\] and Kwoi Tai Chui et al. \[5\] revealed that worldwide great efforts are being made in the field of healthcare, water pollution \[6\] and air pollution \[38\].

### 4.2. Benefits of ML applications for the public sector

Regarding the main benefits of using ML in the public sector, the ML systems are efficient, accurate, with high performance, and usable to different domains, especially for solving classification problems. The quality of the provision of governments to the citizens can be improved by high accuracy in governmental documents \[9\]. Furthermore, ML consists an easier and faster way for automated classification for data analysis, by reducing the cost and the complexity of alternative processes, compared to manual processes which consume a significant amount of effort and time \[6, 7\]. In addition, efforts are majorly devoted to different domains, especially to the domains where the influence is higher to society such as education, migration, energy, urbanization, and healthcare \[3\]. This is also related with another positive aspect of ML, i.e. scalability, as reusability of ML models is considered very high \[8, 10, 11\].

Sentiment Analysis (SA) using ML is a technique capable of enhancing the interaction, and thus the relationship between governments and citizens. SA has the ability to eliminate possible criticalities among the two parts (government and citizens) leading the first on taking the right decisions and actions \[74\]. Considering social media, SA allows governmental organizations to identify and understand citizens’ needs while enabling citizens on affecting either the service delivery or the implementation of any new service or allows them to identify new innovations for services that already exist \[73\]. Table 1 summarizes the benefits of the adoption of ML, analyzed in specific topics related to government services, pinpointing at the research, where these benefits are cited.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Details</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>The pervasiveness of the digital information age further leads to the generation of big data as well as big government data, at a faster rate, thus making manual data analysis and interpretation impossible. ML does not only automate the analysis of big government data but also can provide data-driven answers to vexing questions and even can help in the creation of new theories.</td>
<td>[11]</td>
</tr>
<tr>
<td>Accuracy</td>
<td>The results of ML systems, irrespective of the used techniques, are more accurate, since ML can process big government data and no intervention from either knowledge engineers or domain experts is required.</td>
<td>[8, 9, 10]</td>
</tr>
<tr>
<td>Performance &amp; Process Simplification</td>
<td>ML consists an easier and faster way for automated classification to analyse data when compared to manual process which would consume a significant amount of time and effort by reducing the cost and the complexity of alternative processes.</td>
<td>[6, 7]</td>
</tr>
<tr>
<td>Flexibility</td>
<td>ML efforts are majorly devoted to different domains especially to the more influencing applications to the society such as healthcare, energy crisis, education, food security, overfishing, environmental pollution, migration crisis, urbanization, and water security.</td>
<td>[3, 5]</td>
</tr>
<tr>
<td>Multi-dimensional</td>
<td>Multi-dimensional and multi-variety data can be handled through ML either in dynamic or even in uncertain environments.</td>
<td>[13]</td>
</tr>
<tr>
<td>Team – Based &amp; Mixed-Initiative Learning</td>
<td>New ML methods are capable of collaboratively works with humans analysing complex datasets. Machines could extract information from massive data while humans could suggest new hypotheses and generate explanations based on the information extracted from the machines.</td>
<td>[5]</td>
</tr>
<tr>
<td>Data Utilization</td>
<td>Although a huge amount of government data is already open and online, in many cases, societies, currently, have not the mechanisms, or laws or even cultures to benefit from them. An ML system can be beneficial concerning their utilization and exploitation.</td>
<td>[5]</td>
</tr>
<tr>
<td>Continuous Improvement</td>
<td>ML systems have the capability of the continuous “self-improvement” by using historical data.</td>
<td>[12]</td>
</tr>
<tr>
<td>Scalability</td>
<td>There are techniques that can be used in a ML model in order for it to be upgraded and used for several reasons (such as numeric predictions and decision making).</td>
<td>[5, 8, 10, 11]</td>
</tr>
</tbody>
</table>
4.3. Challenges of using machine learning in government

In the utilization of ML in government there are also limitations which are primarily imposed by the nature of the data analyzed and may lead to misleading results. Both the development of a ML model and the data processing may cost a significant amount of time, considering that, for the former, there are no specific techniques to be followed, while in many cases a variety of techniques have to be tested and, for the latter, the amount of data may be tremendous. Table 2 summarizes the challenges using ML in government services.

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>DETAILS</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy &amp; Ethical Issues</td>
<td>In many cases (e.g., healthcare) the collection of personal data and the ownership of personal data and the benefits of their processing leads to privacy and ethical issues.</td>
<td>[5]</td>
</tr>
<tr>
<td>Combination of Various ML Techniques</td>
<td>In many cases, based on which action to be taken and when to be taken, various ML techniques are needed to be tried for the extraction of better and proper results. So in many cases a significant amount of time may be needed.</td>
<td>[20]</td>
</tr>
<tr>
<td>Legal Issues – GDPR (General Data Protection Regulation)</td>
<td>ML depends upon collecting and processing data from society. This data, in many cases, is explicitly sensitive (e.g., racial origin, religious, health data, ethnic origin).</td>
<td>[61, 62]</td>
</tr>
<tr>
<td>Quality and Quantity of Data</td>
<td>Lack of data (geographical data) or even accessed data may not be representative and, in cases of predictions, barriers can be found decreasing the quality and quantity of the ML system.</td>
<td>[16, 17]</td>
</tr>
<tr>
<td>Unstructured Data</td>
<td>Unstructured data are a major challenge in the usage of ML if we consider the different regional languages.</td>
<td>[16, 17, 21]</td>
</tr>
<tr>
<td>Interpretation of results</td>
<td>Interpretation of results is also a major challenge to determine the effectiveness of ML algorithms. Although an ML algorithm can extract results correctly, their interpretation may not be proper and subject to the human factor.</td>
<td>[18]</td>
</tr>
<tr>
<td>Information Overload</td>
<td>The computational power in combination with big data is less efficiency since ML processes should be capable of removing and neglecting data in order to enjoy a finite and reasonable computation time.</td>
<td>[1, 17, 21]</td>
</tr>
<tr>
<td>Heterogeneity of Data</td>
<td>Heterogeneity of data (e.g., writing styles or different vocabularies) is a challenge for ML because it can lead to wrong results.</td>
<td>[19]</td>
</tr>
<tr>
<td>Availability of Data</td>
<td>In many cases there can be found difficulties of gaining regulatory approval of accessing data (for instance in healthcare), or even lack of data (geographical data) in order for a ML system to be properly trained for quality results.</td>
<td>[14, 15]</td>
</tr>
</tbody>
</table>

The grand challenges that ML usage in government are confronted are related with the quality and quantity of data [16, 17]. If we also considered the existence of different regional languages, heterogeneity and unstructured data [16, 17] raises the complexity of ML tasks. Two contradictive challenges occur; the one is lacking relevant data for processing [14, 15], whereas in some cases too much data is produced leading to information overload problems [17, 3] Data sets may include sensitive data (e.g., health records) whose acquisition and processing is prohibited by third parties without proper permission. Although by removing sensitive data from datasets will solve the legal issues, no one can ensure proper predictions from the ML model without these types of data (e.g., postal code can reveal racial information and loan defaulting) [61].

With regard to legal, ethical barriers, we identified difficulties of gaining regulatory approval of accessing data (for instance, in healthcare or geographical data) or even when limited access is permitted, data may not be representative or incomplete. Unstructured data consist a major challenge in the usage of ML (e.g., if we think of the different regional languages) and in combination with the limited computational power and the Big Data, the efficiency loss may be unavoidable.

5. LANDSCAPING MACHINE LEARNING PROJECTS IN E-GOVERNMENT

This section a detailed presentation of the identified projects, as resulted from the second step of our methodology. It has been revealed that the number of public sector organizations which are being involved to the implementation of ML solutions is continuously growing. Their intended result moves towards the exploitation of the advantages that ML implementations can provide to those services.

Particularly, we identified 18 projects including projects that started from 2011 or later 2018. Generally, there are few cases of ML usage by governments; the most relevant of them are presented on Table 3. In particular, this study is focused on the analysis of the projects from the perspective of their TR Level (Technology Readiness Level), the application areas they cover and the contributed governmental organizations in the projects’ implementation, by including the name of the countries that these governmental organizations represent. The analysis of the projects based on the three abovementioned criteria, results in the following observations:

4) Considering the Technology Readiness Level, the majority of the projects fall into the second category among the three different groups in which they were categorised: (a) proof of concept, (b) pilot applications and (c) large-scale implementations.

5) Flexibility as one of the key benefits of the ML applications identified in the previous step of the methodology is verified, since a multitude of application domains have been revealed from the analysis. The diverse sectors covered by the projects are as follows: Health, Energy, Transport, Society, Climate, ICT Research & Innovation, Entrepreneurship, Culture, Social
In all projects, there is a cooperation among governmental and private organizations for their implementation and in most of the cases, governmental organizations contribute as pilots.

Although the majority of ML initiatives, as we mentioned above, are being implemented with the cooperation among public and private sector, the solutions can be beneficial for both sectors.

Until now, all ML initiatives are mainly used as supporting systems for data-driven decision-making solutions.

In most of the cases, two or more governmental organizations from different countries cooperated for the implementation of an ML initiative.

Table 3: Machine Learning projects in e-Government

<table>
<thead>
<tr>
<th>Project Title</th>
<th>TRL Level</th>
<th>Application Areas</th>
<th>Government</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Policy Mashing of Gadgets Underlying Knowledge in Web 2.0 Media (PADGETS)</td>
<td>Pilot</td>
<td>Energy, Entrepreneurship, Health, Finance</td>
<td>Greece, Slovenia, Italy</td>
<td>Design, develop and deploy a prototype tool-set to allow policy makers to graphically create web applications, deployed in Web 2.0 media.</td>
</tr>
<tr>
<td>EU Community</td>
<td>Pilot</td>
<td>Energy, Entrepreneurship &amp; Innovation, Future of Europe</td>
<td>Europea</td>
<td>Combination of social media interactions, qualified contributors, document curation, visual analysis plus online and offline trust-building tools for the provision of better policy options to decision makers.</td>
</tr>
<tr>
<td>Policy Formulation and Validation through non-moderated crowdsourcing (NOMAD)</td>
<td>Pilot</td>
<td>Energy, Health, Transparency</td>
<td>Greece, Austria, United Kingdom</td>
<td>Introduction of new dimensions into the experience of policy making by providing decision-makers with fully automated solutions for content search, selection acquisition, categorization and visualization that work in a collaborative form in the policy-making arena.</td>
</tr>
<tr>
<td>EU-wide Legal Text Mining</td>
<td>Pilot</td>
<td>Legislation</td>
<td>Greece, Austria</td>
<td>Seamless and inclusive access to legal information</td>
</tr>
</tbody>
</table>

Using Big Data Processing Infrastructure, delivering Advanced Services for Citizens, Businesses and Administrations (MANYLAW)

<table>
<thead>
<tr>
<th>Recognition and Enrichment of Archival Documents (READ)</th>
<th>Large Scale Implementation</th>
<th>Culture (Historical, Handwritten documents)</th>
<th>Greece, Finland, Germany, Switzerland</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIMplifying the interaction with Public Administration Through Information Technology for Citizens and Companies (SIMPATICO)</td>
<td>Pilot</td>
<td>Society (e-services)</td>
<td>Italy, United Kingdom, Spain</td>
</tr>
<tr>
<td>CORE: the digital showcase for Europe</td>
<td>Pilot</td>
<td></td>
<td>United Kingdom, Greece, Italy, Finland, Holland, Denmark, Norway, Spain</td>
</tr>
<tr>
<td>Development, deployment and operation of the technical and organisational foundations of an open, multi-vendor platform for Europe-wide mobility services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOBI-NET: the EU-wide e-market place of mobility services for businesses and end users</td>
<td>large scale implementation</td>
<td>Transport</td>
<td>United Kingdom, Greece, Italy, Finland, Holland, Denmark, Norway, Spain</td>
</tr>
<tr>
<td>PERICLES: Promoting and Enhancing Re-use of</td>
<td>Proof of Concept</td>
<td>Culture</td>
<td>United Kingdom, Belgium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information throughout the Current Lifecycle taking account of Evolving Semantics (Pericles)</td>
<td>remains accessible in an environment that is subject to continual change.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EURECA</td>
<td>Large-Scale Implementation</td>
<td>ICT Research and Innovation, Health</td>
<td>United Kingdom, The European Institute for Innovation Through Health Data</td>
</tr>
<tr>
<td>INFRALETR: LINEAR INFRASTRUCTURE EFFICIENCY IMPROVEMENT BY AUTOMATED LEARNING AND OPTIMISED PREDICTIVE MAINTENANCE TECHNOQUIS (INFRALETR)</td>
<td>Pilot</td>
<td>Transport (Road Network, Railroad Network)</td>
<td>Portugal, Sweden</td>
</tr>
<tr>
<td>X5GON: Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network: Artificial Intelligence and Open Educational Re-sources (X5GON)</td>
<td>Pilot</td>
<td>Education</td>
<td>Slovenia</td>
</tr>
<tr>
<td>STREAMLINE</td>
<td>Pilot</td>
<td>Social Sciences and Humanities, Society</td>
<td>Germany, Hungary</td>
</tr>
<tr>
<td>A Holistic, Innovative Framework for Design, Development and Orchestration of 5G-ready Applications and Network Services over Sliced Programmable Infrastructure (MATILDA)</td>
<td>Pilot</td>
<td>ICT Research and Innovation</td>
<td>Greece</td>
</tr>
<tr>
<td>Big Data Europe: Empowering Communities with Data Technologies (BIG DATA EUROPE)</td>
<td>Large Scale Implementation</td>
<td>Health, Food and Agriculture, Energy, Transport, Climate, Social Sciences, Security</td>
<td>Greece, Italy, Norway, Spain</td>
</tr>
<tr>
<td>Holistic Benchmarking of Big Linked Data (HOB-BIT)</td>
<td>Large Scale Implementation</td>
<td>Industry, Environment (Geospatial Data Analysis), Energy (Smart Energy), Climate (Weather Data Analysis), Society (Human Resource Management, European Societal Challenges), ICT Research &amp; Innovation (Enterprise Research)</td>
<td>Greece</td>
</tr>
</tbody>
</table>
| collaborati ve Management Platform for detection and Analyses of (Re-) emerging and foodborne out-breaks in Europe | Large Scale Implementation | Health, Food (emerging epide miologies research in the human, animal and food sectors) | Denmark, France, Italy, Netherlands, UK, Hungary | Integration of state-of-the-art strategies, tools, techniques and methods for collecting, processing and analysing sequence-based pathogen data in combination with associated (clinical, epidemiological and other)
6. CONCLUSIONS

The scope of this study was the identification of benefits and obstacles towards the adoption of the ML innovative technology and the identification of ML approaches in the public sector. To address this, a thorough review on specific databases has been conducted. Our findings indicate that ML can be used to analyze Big Data, including government’s data, or even to generate new knowledge, while classification problems can be met. Despite the different available approaches for solving classification problems, the main approach is ML techniques due to the capabilities that these can provide. From the conducted analysis, it is revealed that ML is a method used in order to devise complex systems and, by using statistical techniques, it consists a powerful tool which can predict or support governments’ decision makers. Among the most cited perks of ML usage in government is accuracy, efficiency, scalability and flexibility.

However, the benefits that ML can provide to governments are possible limited because of the nature of data and the human intervention needed for the interpretation of results, which may lead to misleading results. For the integration of different datasets categories [16] ML can create new ways to solve complex problems (sarcasm) by combining, also, deep learning techniques with neural network [6] to avoid possible difficulties concerning computing power and time issues. Despite the barriers concerned by the nature of data, ML is proven to be a promising technology and the public sector can not be excluded estimating, also, all the potential benefits provided by the ML. Detailed datasets (such us crime and health statistics) consist extremely useful information for decision makers and can be extracted through the ML. ML models’ effectiveness can be improved by using and testing a variety of techniques and there are also many available free tools implemented for that reason [72]. Furthermore, as Alexandra Terlyga & Igor Balk [71] mentioned that governments can use ML clustering techniques to set goals to their units, based on different indicators (e.g., clustering high-level educational institutions by overall spending (Profit and Loss Statement), can lead governments on setting proper expectations.

Our study, also, revealed that a variety of applications in different domains is being implemented with all governmental organizations participating as pilots. Although a large port of the applications executed only as pilots, it is clear enough that many governments not only consider on improving their existing services, but actions are conducted towards decision making through the use of ML. Analysis of these projects, also identified the need for collaboration with private sector organizations.

One of the aspects to which further research should target is the development of frameworks for addressing the challenges of Big Data ML, such as the data’s heterogeneity and the integration of datasets from different domains. Similar efforts al-ready exist and can lead governments to more accurate results for data-driven decision making [76]. Another stream of future research is to create a legal frame on how these data can be used for the common good, overcoming the limitations of the usage of datasets which include sensitive information, imposed by the recent EC regulation for the protection of individuals privacy [75].

ACKNOWLEDGMENTS

The research leading to these results is developed in the context of the Gov 3.0 Project. It has received funding from the Erasmus+ Knowledge Alliance, Project Reference No. 588306-EPP-1-2017-1-EL-EPPKA2-KA.

REFERENCES

How Machine Learning is Changing e-Government

ICEGOV2019, 3-5 April 2019, Melbourne, VIC, Australia


of such data, and repealing directive 95/46/EC (General Data Protection
Regulation) (2016)

