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
vdiamant@aegean.gr

Yannis Charalabidis
University of the Aegean
81100, Mytilene
Greece
yannisx@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.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ICEGOV2019, April 3–5, 2019, Melbourne, VIC, Australia
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6644-1/19/04…$15.00
https://doi.org/10.1145/3326365.3326412

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:

C. Alexopoulos, Z. Lachana, A. Androutsopoulou, V. Diamantopoulou, Y. Charalabidis, M. Avgerinos Loutsaris. 2019. How Machine Learning is Changing e-Government. In *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance (ICEGOV2019), Melbourne, VIC, Australia, April 3-5, 2019*, 10 pages. https://doi.org/10.1145/3326365.3326412

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- net 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.tc. (allowing users to create content and amplify also the huge web volume) [23]. 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 peek 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 Ma- chine 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 pub- lic 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 hu- mans 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 Kwok Tai Chui et al. [3] 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 interactivity, 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 1: Potential Benefits of ML

TOPIC	DETAILS	REFERENCES
Efficiency	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.	[11]
Accuracy	The results of ML systems, irrespec- tive 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.	[8, 9, 10]
Performance & Process Simplification	ML consists an easier and faster way for automated classification to analyse data when compared to manual process which would con- sume a significant amount of time and effort by reducing the cost and the complexity of alternative pro- cesses.	[6, 7]
Flexibility	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 pollu- tion, migration crisis, urbanization, and water security.	[3, 5]
Multi-dimensional	Multi-dimensional and multi-variety data can be handled through ML ei- ther in dynamic or even in uncertain environments.	[13]
Team – Based & Mixed-Initiative Learning	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.	[5]
Data Utilization	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	[5]
Continuous Improvement	ML systems have the capability of the continuous "self-improvement" by using historical data.	[12]
Scalability	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)	[3, 8, 10, 11]

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 2: Potential Challenges of ML

TOPIC	DETAILS	REFERENCES
Privacy & Ethical Issues	In many cases (e.g. healthcare) the collection of personal data, the ownership of personal data and the benefits of their processing leads to privacy and ethical issues	[5]
Combination of Various ML Techniques	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.	[20]
Legal Issues – GDPR (General Data Protection Regulation)	ML depends upon collecting and processing data from society. This data, in many cases, is explicitly sensitive (e.g., racial origin, reli- gious, health data, ethnic origin).	[61, 62]
Quality and Quantity of Data	Lack of data (geographical data) or even accessed data may not be rep- resentative and, in cases of predic- tions, barriers can be found decreas- ing the quality and quantity of the ML system.	[16, 17]
Unstructured Data	Unstructured data are a major chal- lenge in the usage of ML if we con- sider the different regional lan- guages.	[16, 1,7, 21]
Interpretation of results	Interpretation of results is also a ma- jor challenge to determine the effec- tiveness of ML algorithms. Alt- hough an ML algorithm can extract results correctly, their interpretation may not be proper and subject to the human factor.	[18]
Information Overload	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.	[3, 17, 21]
Heterogeneity of Data	Heterogeneity of data (e.g. writing styles or different vocabularies) is a challenge for ML because it can lead to wrong results	[19]
Availability of Data	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.	[14, 15]

The grand challenges that ML usage in government are con- fronted 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 oc- cur; 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 pro- cessing is prohibited by third parties without proper permis- sion. 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 difficul- ties of gaining regulatory approval of accessing data (for in- stance, in healthcare or geographical data) or even when lim- ited access is permitted, data may not be representative or in- complete. 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 to- wards 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 do- mains have been revealed from the analysis. The di- verse sectors covered by the projects are as follows: Health, Energy, Transport, Society, Climate, ICT Re- search & Innovation, Entrepreneurship, Culture, Social

Sciences & Humanities, Industry, Environment, Food and Agriculture, Finance, Future of Europe, Transparency, Legislation, Education, Security.

6) 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.
7) 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.
8) Until now, all ML initiatives are mainly used as supporting systems for data-driven decision-making solutions.
9) 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

Project Title	TRL Level	Application Areas	Govern ments	Scope
Policy Gadgets Mashing Underlying Group Knowledge in Web 2.0 Media (PADGETS)	Pilot	Energy, Entrepreneur ship, Health, Finance	Greece, Slovenia , Italy	Design, develop and deploy a prototype tool- set to allow pol- icy makers to graphically create web applications, deployed in Web 2.0 media.
EU Community	Pilot	Energy, Entrepreneur ship & Innovation, Future of Europe	Europea n Union	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.
Policy Formulation and Validation through non moderated crowdsourci ng (NOMAD)	Pilot	Energy, Health, Transparency	Greece, Austria, United Kingdo m	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.
EU-wide Le- gal Text Mining using Big Data Processing Infra- structures, delivering Advanced Ser- vices for Citizens, Businesses and Administrati ons MANYLAW S	Pilot	Legislation	Greece, Austria	Seamless and inclusive access to legal information across EU and improve the efficacy of decision making in legislative procedures operated by governmental organizations.
Recognition and Enrichment of Archival Documents (READ)	Large Scale Implemen tation	Culture (Historical, Handwritten documents)	Greece, Finland, Ger- many, Switzer- land	Implementation of a Virtual Re- search Environment where archivists, humanities scholars, computer scientists and volunteers can boost research, innovation, development and usage of cutting edge technology for the automated recognition ,transcription, indexing and enrichment of handwritten archival documents.
SIMplifying the interaction with Public Administrati on Through Information technology for Citizens and cOmpanies - SIMPATICO	Pilot	Society (e- ser- vices)	Italy, United Kingdo m, Spain	SIMPATICO's "learning by doing" approach will use this information and match it with user pro- files to continuously adapt and improve interactions with the public services. All the collected information on public services and procedures will be made available within Citizenpedia, a collective knowledge database released as a new public domain re- source.
MOBiNET: the Eu- wide e-market place of mobility services for businesses and end users /MOBiNET	large scale implemen tation	Transport	United Kingdo m, Greece, Italy, Finland, Holland, Den- mark, Norway, Spain	Development, deployment and operation of the technical and organisational foundations of an open, multi- vendor platform for Europe- wide mobility services)
Pericles: Promoting and Enhancing Re- use of	Proof of Concept	Culture	United Kingdo m, Belgium	PERICLES aims to address the challenge of ensuring that digital content

Information throughout the Content Lifecycle taking account of Evolving Semantics (Pericles)				remains accessible in an environment that is subject to continual change.
EURECA	Large-Scale Implementation	ICT Research and Innovation, Health	United Kingdom, The European Institute For Innovation Through Health Data	EURECA aims to build soft- ware solutions to improve interoperability among existing data systems,such as clinical trials and electronic health record systems..
INFRALERT: LINEAR INFRASTRUCTURE EFFICIENCY IMPROVEMENT BY AUTOMATED LEARNING AND OPTIMISED PREDICTIVE MAINTENANCE TECHNIQUES (INFRALERT)	Pilot	Transport (Road Network, Railroad Network)	Portugal, Sweden	Development of an expert-based information system to support and auto- mate infrastructure management from measurement to maintenance.
X5GON: Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global OER Network- Artificial Intelligence and Open Educational Re- sources (X5GON)	Pilot	Education	Slovenia	Solution helping users/students find what they need not just in OER repositories, but across all open educational re- sources on the web.
STREAM-LINE	Pilot	Social Sciences and Humanities, Society	Ger-many, Hungary	Addresses the competitive ad- vantage needs of European online mediabusinesses(EOMB) by delivering fast re- active analytics suitable in solving a wide array of problems, including ad- dressing customer retention, personalised recommendation,

				and more broadly targeted services.
A Holistic, Innovative Framework for Design, Development and Orchestration of 5G-ready Applications and Network Services over Sliced Programmable Infrastructure /MATILDA	Pilot	ICT Research and Innovation	Greece	Design and implementation of a holistic 5G end-to-end services operational frame-work tackling the lifecycle of design, development and orchestration of 5G-ready applications and 5G network ser-vices over programmable infrastructure, following a unified programmability model and a set of control abstractions
Big Data Europe – Empowering Communities with Data Technologies (BIG DATA EUROPE)	Large Scale Implementation	Health, Food and Agriculture, Energy, Transport, Climate, Social Sciences, Security	Greece, Italy, Norway, Spain	BDE aims to build an extensive stakeholder network spread within relevant communities from across the different SC domains; cover the whole process of data us- age within each, from data collection,processing, storage and visualization to the development of data services.
Holistic Benchmarking of Big Linked Data (HOB- BIT)	Large Scale Implementation	Industry, Environment (Geospatial Data Analysis), Energy (Smart Energy), Climate (Weather Data Analysis), Society(Human Resource Management, European Societal Challenges), ICT Research & Innovation (Enterprise Research)	Greece	HOBBIT aims at abolishing the barriers in the adoption and deployment of Big Linked Data by European companies, by means of open bench- marking reports that allow them to assess the fit- ness of existing solutions for their purposes
COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne out- breaks in Eu- rope	Large Scale Implementation	Health, Food (emerging epidemics research in the human, animal and food sectors)	Den-mark, France, Italy, Nether-lands, UK, Hungary	Integration of state-of-the-art strategies, tools, technologies and methods for collecting, processing and analysing sequence-based pathogen data in combination with associated (clinical,epidemiological and other)

(COM-PARE)				data, for the generation of actionable information to relevant authorities and other users in the human health, animal health and food safety domains.
Cyber Physical System based Proactive Collaborative Maintenance (MAN-TIS)	Large Scale Implementation	Industry Production asset maintenance, Transport-Vehicle maintenance management, Energy(Energy production asset management),Health (Health equipment maintenance)	Austria, Germany	Development of a Cyber Physical System based on Pro- active Maintenance Service Platform Architecture enabling Collaborative Maintenance Ecosystems.
SWAMI: Space Weather Atmosphere Models and Indices (SWAMI)	proof of concept	Space, Climate, Environment	France, United Kingdom, Ger-many	Provision of an improved and comprehensiverepresentation of the neutral atmosphere from the sur- face to 1500 km altitude.

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 sec- tor. 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 de- vise 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 us- age 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 hu- man 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 (sarcastic remarks) 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 pro- vided 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 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. 588306EPP-1-2017-1-EL-EPPKA2-KA.

REFERENCES

[1] Pereira, G. V., Charalabidis, Y., Alexopoulos, C., Mureddu, F., Parycek, P., Ronzhyn, A., ... & Wimmer, M. A. (2018, May). Scientific foundations training and entrepreneurship activities in the domain of ICT-enabled governance. In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age (p. 98). ACM.
[2] Lachana, Z., Alexopoulos, C., Loukis, E., Charalabidis, Y. (2018, September). IDENTIFYING THE DIFFERENT GENERATIONS OF EGOVERNMENT: AN ANALYSIS FRAMEWORK
[3] Chui, K. T., Alhalabi, W., Pang, S. S. H., Pablos, P. O. D., Liu, R. W., & Zhao, M. (2017). Disease Diagnosis in Smart Healthcare: Innovation, Technologies and Applications. Sustainability, 9(12), 2309. https://doi.org/10.3390/su9122309
[4] Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine, 23(1), 89-109.
[5] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.
[6] Patil, A. P., Doshi, D., Dalsaniya, D., & Rashmi, B. S. (2017, September). Applying Machine Learning Techniques for Sentiment Analysis in the Case Study of Indian Politics. In International Symposium on Signal Processing and Intelligent Recognition Systems (pp. 351-358). Springer, Cham. https://doi.org/10.1007/978-3-319-67934-1_31
[7] Marasco, D. E., & Kontokosta, C. E. (2016). Applications of machine learning methods to identifying and predicting building retrofit opportunities. Energy and Buildings, 128, 431-441. https://doi.org/10.1016/j.enbuild.2016.06.092

[8] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.
[9] Moldagulova, A., & Sulaiman, R. B. (2017, May). Using KNN algorithm for classification of textual documents. In Information Technology (ICIT), 2017 8th International Conference on (pp. 665-671). IEEE.
[10] Mitchell, T. M. (2006). The discipline of machine learning (Vol. 9). Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Machine Learning Department.
[11] Grimmer, J. (2015). We are all social scientists now: how big data, machine learning, and causal inference work together. PS: Political Science & Politics, 48(1), 80-83.
[12] Cios, K. J., Pedrycz, W., Swiniarski, R., & Kurgan, L. A. 2007. "Supervised learning: Decision trees, rule algorithms, and their hybrids." Data mining: A knowledge discovery approach. NY: Springer Science and Business Media
[13] Marsland, S. (2011). Machine learning: an algorithmic perspective. Chapman and Hall/CRC.
[14] Wilbanks, J. T., & Topol, E. J. (2016). Stop the privatization of health data. Nature News, 535(7612), 345. https://doi.org/10.1038/535345a
[15] Piscopo, A., Siebes, R., & Hardman, L. (2017). Predicting sense of community and participation by applying machine learning to open government data. Policy & Internet, 9(1), 55-75. https://doi.org/10.1002/poi3.145
[16] Eshleman, R., & Yang, H. (2014, December). " Hey# 311, Come Clean My Street!": A Spatio-temporal Sentiment Analysis of Twitter Data and 311 Civil Complaints. In Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on (pp. 477-484). IEEE. https://doi.org/10.1109/BDCloud.2014.106
[17] Netten, N., van den Braak, S., Choenni, S., & van Someren, M. (2016, March). A Big Data Approach to Support Information Distribution in Crisis Response. In Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance (pp. 266-275). ACM. https://doi.org/10.1145/2910019.2910033
[18] Sommer, R., & Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In Security and Privacy (SP), 2010 IEEE Symposium on (pp. 305-316). IEEE.
[19] Torii, M., Wagholikar, K., & Liu, H. (2011). Using machine learning for concept extraction on clinical documents from multiple data sources. Journal of the American Medical Informatics Association, 18(5), 580-587.
[20] RF Wireless World, 2012Advantages of Machine Learning | Disadvantages of Machine Learning. Retrieved from http://www.rfwireless-world.com
[21] Suthaharan, S. (2014). Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Performance Evaluation Review, 41(4), 70-73.
[22] S. M. Weiss and N. Indurkhya. Predictive data mining: a practical guide. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.
[23] Fan, W., & Bifet, A. (2013). Mining big data: current status, and forecast to the future. ACM sIGKDD Explorations Newsletter, 14(2), 1-5.
[24] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp. 42-47). IEEE.)
[25] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,.
[26] F. Diebold. On the Origin(s) and Development of the Term "Big Data". Pier working paper archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012
[27]] Abbod, M. F., Catto, J. W., Linkens, D. A., & Hamdy, F. C. (2007). Application of artificial intelligence to the management of urological cancer. The Journal of Urology, 178(4), 1150-1156
[28] Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., & Trevisan, D. (2017, June). Transparency in practice: using visualization to enhance the interpretability of open data. In Proceedings of the 18th Annual International Conference on Digital Government Research (pp. 139-148). ACM. https://doi.org/10.1145/3085228.3085294
[29] Huang, H. H., & Liu, H. (2014, October). Big data machine learning and graph analytics: Current state and future challenges. In Big Data (Big Data), 2014 IEEE International Conference on (pp. 16-17). IEEE. https://doi.org/10.1109/BigData.2014.7004471
[30] Kelly, J., Floyer, D., Vellante, D., & Miniman, S. (2014). Big data vendor revenue and market forecast 2012-2017. https://wikibon.com/big-data-vendor-revenue-and-market-forecast-2011-2026/
[31] Yunhong Zhou, Dennis Wilkinson, Robert Schreiber, and Rong Pan. Large-scale parallel collaborative filtering for the netflix prize. In Algorithmic Aspects in Information and Management, pages 337–348. Springer, 2008. https://link.springer.com/chapter/10.1007/978-3-540-68880-8_32
[32] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of machine Learning research, 2003. http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf
[33] Amol Ghoting, Rajasekar Krishnamurthy, Edwin Pednault, Berthold Reinwald, Vikas Sindhwani, Shirish Tatikonda, Yuanyuan Tian, and Shivakumar Vaithyanathan. Systemml: Declarative machine learning on mapreduce. In IEEE International Conference on Data Engineering (ICDE), 2011, pages 231–242. http://ieeexplore.ieee.org/abstract/document/5767930/
[34] Yucheng Low, Danny Bickson, Joseph Gonzalez, Carlos Guestrin, Aapo Kyrola, and Joseph M Hellerstein. Distributed graphlab: a framework for machine learning and data mining in the cloud. Proceedings of the VLDB Endowment (VLDB), 2012. http://vldb.org/pvldb/vol5/p716_yuchenglow_vldb2012.pdf
[35] Toshihiko Sakai and Sachio Hirokawa. 2012. Feature words that classify problem sentence in scientific article. In the 14th International Conference on Information Integration and Web-based Applications & Services. ACM, 360–367. https://dl.acm.org/citation.cfm?id=2428803
[36] Hirokawa, S., Suzuki, T., & Mine, T. (2017, August). Machine learning is better than human to satisfy decision by majority. In Proceedings of the International Conference on Web Intelligence (pp. 694-701). ACM. ISO 690 https://doi.org/10.1145/3106426.3106520
[37] Ghosh, D., Chun, S., Shafiq, B., & Adam, N. R. (2016, June). Big Data-based Smart City Platform: Real-Time Crime Analysis. In Proceedings of the 17th International Digital Government Research Conference on Digital Government Research (pp. 58-66). ACM. https://doi.org/10.1145/2912160.2912205
[38] Hu, K., Rahman, A., Bhrugubanda, H., & Sivaraman, V. (2017). HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors. IEEE Sensors Journal, 17(11), 3517-3525. https://doi.org/10.1109/JSEN.2017.2690975
[39] Tomašev, N. (2017). Extracting the patterns of truthfulness from political information systems in Serbia. Information Systems Frontiers, 19(1), 109-127. https://doi.org/10.1007/s10796-015-9596-8
[40] Hollings, T., Robinson, A., van Andel, M., Jewell, C., & Burgman, M. (2017). Species distribution models: A comparison of statistical approaches for livestock and disease epidemics. PloS one, 12(8), e0183626. https://doi.org/10.1371/journal.pone.0183626
[41] Chang, Y. W., & Tsai, C. Y. (2017, March). Apply Deep Learning Neural Network to Forecast Number of Tourists. In Advanced Information Networking and Applications Workshops (WAINA), 2017 31st International Conference on (pp. 259-264). IEEE. https://doi.org/10.1109/WAINA.2017.125
[42] Amma, N. B. (2016). Big data mining. In Effective Big Data Management and Opportunities for Implementation (pp. 53-59). IGI Global.
[43] Manovich, L. (2011) 'Trending: the promises and the challenges of big social data', in Debates in the Digital Humanities, ed. M. K. Gold, The University of Minnesota Press, Minneapolis, MN, [Online] Available at: http://www.manovich. net/DOCS/Manovich_trending_paper.pdf (15 July 2011).
[44] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
[45] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.
[46] Cuzzocrea, A., Song, I. Y., & Davis, K. C. (2011, October). Analytics over large-scale multidimensional data: the big data revolution!. In Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP (pp. 101-104). ACM.
[47] Elgendy, N., & Elragal, A. (2014, July). Big data analytics: a literature review paper. In Industrial Conference on Data Mining(pp. 214-227). Springer, Cham.
[48] Charalabidis, Y. The World of Open Data: Concepts, Methods, Tools and Experiences. Springer
[49] Gartner, http://www.gartner.com/it-glossary/bigdata
[50] McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review, 90, 60–66, 68, 128.
[51] D. Laney. 3-D Data Management: Controlling Data Volume, Velocity and Variety. META Group Research Note, February 6, 2001
[52] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). Machine learning: An artificial intelligence approach. Springer Science & Business Media.
[53] Jeffcock Peter (2017) https://blogs.oracle.com/bigdata/machine-learning-techniques
[54] Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135). Cambridge: MIT press.
[55] Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. Econometric Reviews, 29(5-6), 594-621.

[56] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.
[57] Eisenecker, U. W. (1995). Ai: The tumultuous history of the search for artificial intelligence. AI Communications, 8(1), 45-47.
[58] Buchanan, B. G. (2005). A (very) brief history of artificial intelligence. Ai Magazine, 26(4), 53.
[59] Nilsson, N. J. (2014). Principles of artificial intelligence. Morgan Kaufmann.
[60] Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. Harv. JL & Tech., 29, 353.
[61] Goodman, B., & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a" right to explanation". arXiv preprint arXiv:1606.08813.
[62] Calders, T., and Verwer, S. 2010. Three Naive Bayes Approaches for DiscriminationFree Classification. Data Mining and Knowledge Discovery 21(2): 277–292. doi.org/ 10.1007/s10618-010-0190-x
[63] Buzzle. (n.d.). Unbelievably Brilliant Applications of Artificial Intelligence. Retrieved January 8, 2018, from https://www.buzzle.com/articles/applications-of-artificial-intelligence.html Eggers. William, Schatsky, D., & Viechnicki, P. (2017). AI-augment
[64] Business Insider. (2016). The ATO's virtual assistant has already answered almost a million enquiries this year | Business Insider. Retrieved January 8, 2018, from https://www.businessinsider.com.au/the-ato-launched-a-siri-for-tax-and-has-called-it-alex-2016-12
Capgemini. (2017). Unleashing the potential of Artificial Intelligence in the Public Sector. Retrieved from https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf
[65] Mehr, H. (2017). Artificial Intelligence for Citizen Services and Government Artificial Intelligence for Citizen Services and Government artificial intelligence for citizen services and government. Retrieved from https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf
[66] Capgemini. (2017). Unleashing the potential of Artificial Intelligence in the Public Sector. Retrieved from https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/10/ai-in-public-sector.pdf
[67] Eggers. William, Schatsky, D., & Viechnicki, P. (2017). AI-augmented government Using cognitive technologies to redesign public sector work. Deloitte University Press. Retrieved from https://www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf
[68] White House, E. O. of the P. (2016). Artificial Intelligence, Automation, and the Economy. Retrieved from https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED
[69] Guszcza, J., "Analyzing Analytics", Contingencies, American Academy of Actuaries, July-August, (2008).].
[70] Zhou, Z. H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives [discussion forum]. IEEE Computational Intelligence Magazine, 9(4), 62-74.
[71] Terlyga, A., & Balk, I. (2017, October). Use of machine learning methods to classify Universities based on the income structure. In Journal of Physics: Conference Series (Vol. 913, No. 1, p. 012005). IOP Publishing. https://doi.org/10.1088/1742-6596/913/1/012005
[72] Nepali, R. K., & Wang, Y. (2016, January). You look suspicious!!: Leveraging visible attributes to classify malicious short urls on twitter. In System Sciences (HICSS), 2016 49th Hawaii International Conference on (pp. 2648-2655). IEEE. https://doi.org/10.1109/HICSS.2016.332
[73] Dolicanin, C., Kajan, E., Randjelovic, D.: Handbook of Research on Democratic Strategies and Citizen-Centered E-Government Services, pp. 231–249. IGI Global, Hersey (2014) https://www.researchgate.net/profile/Nuno_Garcia/publication/290950006_E health_Current_status_and_future_trends/links/56c45c2308aea564e304a9ea.pdf
[74] Corallo, A., Fortunato, L., Matera, M., Alessi, M., Camillò, A., Chetta, V., ... & Storelli, D. (2015, July). Sentiment Analysis for Government: an optimized approach. In International Workshop on Machine Learning and Data Mining in Pattern Recognition (pp. 98-112). Springer, Cham https://doi.org/10.1007/978-3-319-21024-7_7
[75] European parliament: Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation) (2016)
[76] Androutsopoulou, A., & Charalabidis, Y. (2018, April). A framework for evidence based policy making combining big data, dynamic modelling and machine intelligence. In Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance (pp. 575-583). ACM.
[77] Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., & Crombecq, K. (2010). A surrogate modeling and adaptive sampling toolbox for computer based design. Journal of Machine Learning Research, 11(Jul), 2051-2055.
[78] Leonard M., 2018 "Government leans into machine learning", GCN | Technology, Tools and Tactics for Public Sector IT, https://gcn.com/articles/2018/08/17/machine-learning.aspx]