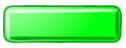
Roadmap for machine learning

Description and state of the art	
Definition	Machine learning is a term that refers to a set of technologies that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. It is closely related to (and often overlaps with) computational statistics, while it has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. *Machine learning* is the subfield of computer science that "gives computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959)[175]. It explores the study and construction of algorithms that can learn from and make predictions on data. Within the field of data analytics in particular, machine learning is a method used to devise complex algorithms that lend themselves to prediction. Such algorithms are composed of many approaches in machine learning, such as deep learning, neural networks and natural-language processing, used in unsupervised and supervised learning that operate guided by lessons from existing information[162]. Originally, targeting to achieve artificial intelligence, machine learning has shifted its focus towards tackling solvable problems of practical nature, whereas it has benefited from the increasing availability of digitized information, and the possibility to distribute that via the Internet[176].
Addressed societal /business or public sector need	Societal need: Inclusive well-being and health
Existing solutions /applications /services	The following solutions are available for implementing machine learning applications: • IBM's Machine Learning[105] • Google AI[177] • Microsoft Azure Machine Learning[178] • Apache Mahout[179] • AmazonML (Amazon Machine Learning)[180] • BigML[181] • Google Prediction API, a Machine Learning black box for

	devs[182] • Wise, Machine Learning for Customer Success[183]
 Main actors regarding R&D of this technology	• IBM • Google • Apache Foundation • Imperial College of Science, Technology and Medicine • Universitat Politecnica de Catalunya • University of Edinburgh • University of Oxford • Institut National de Recherche en Informatique et en Automatique
 Current research activities	Indicative R&D projects include: • **MLPM** ("Machine Learning for Personalized Medicine"), with the goal to educate interdisciplinary experts who will develop and employ the computational and statistical tools that are necessary to enable personalized medical treatment of patients according to their genetic and molecular properties and who are aware of the scientific, clinical and industrial implications of this research[184]. • **SACCSCAN-IA-ML** ("Developing Machine Learning Classifier Models for Eye Movements to Diagnose Major Psychiatric Disorders"), on the development of SaccScan, a novel point-of-care (PoC) software diagnostic system which has been demonstrated to detect schizophrenia with better than 95% accuracy and can be extended with the same precision to other major psychiatric conditions[185]. • **DecoMP_ECoG** ("Decoding memory processing from experimental and spontaneous human brain activity using intracranial electrophysiological recordings and machine learning based methods"), a project to use intracranial electrophysiological recordings from the surface of the human brain to investigate encoding, retrieval and consolidation of category-specific information during experimental settings, as well as during spontaneous brain activity[186]. • **HF-PREDICT**, on the development and validation of the first clinically accurate wearable device and machine learning software for predicting Heart Failure (HF) of a patient[187]. • **HealthSCOPE**, on the delivery of a healthcare scheduling and management system which will enable hospitals to schedule the use of operating theatres, labs and other facilities, allocate staff, select the required equipment and consumables, and allocate bed space for recovery based on the use of cutting-edge machine learning techniques[188].

 Impact assessment	**Public sector modernization:** • Degree of Resources (Capital, Personnel, Infrastructure) Utilization • Efficiency / Productivity • Quality of Services Provided **Public Sector as an Innovation Driver:** • Productivity • Public Safety • Transport Infrastructure • e-Security
Necessary technological modifications	
 Potential use cases	Machine learning systems can be used in the waiting room of a general practitioner to ask the patient about his/her symptoms and suggest the doctor a first diagnose on which the doctor can agree or disagree. DNA sequencing, as well as health data from large pool of users could be used to diagnose diseases and possible health issues, resulting into new studies and more evident based treatment theories.
 Technological challenges	Technological challenges concern the availability and reliability of data, upon which machine learning applications are to be trained. Moreover, as data becomes big data, new algorithms and computational methods are necessary to accelerate the production of results, in acceptable times
Necessary activities (in or for the public sector)	

 Development of a specific training necessary	**Open task**	Users do need to be trained in order for machine learning applications to produce reliable results, both on the mathematical/algorithmic level, as well as data engineering levels.
 Advanced or adapted ICT infrastructure needed	**Open task**	Need for Big Data infrastructure.
 Change of (public sector internal)		No change of public sector internal processes is necessary.

processes necessary		
 Promotion / information of stakeholders necessary	**Open task**	There is a need to promote the advantages of Machine Learning alongside with its precondition for accessing and processing large numbers of data, to allow stakeholders to trust these data intense processes.
 Need to deal with cyber security issues		No cyber security issues identified.
 New or modified legislative framework or regulations necessary	**Open task**	Regulations concerning the use of anonymised personal data would be needed to exploit the full of this technology.
 Development of a common standard necessary		No standards' development is necessary.
 Need for a more economical solution		No need for a more economical solution identified.
Dealing with challenges		
 Ethical issues	**Open task**	Ethical issues may rise a result of the fact that systems which are trained on datasets, collected with biases may exhibit these biases upon use, thus digitizing cultural prejudices such as institutional racism and classism.

 Societal issues	Open task	Concerns may rise around the greater dependence upon technology and the fewer requirements in human resources. Furthermore, decisions proposed by Machine Learning technology are greatly technocratic, and don't take into account societal impact.
 Health issues		No health issues identified.
 Public acceptance		The technology is indeed likely to encounter problems regarding public acceptance, as a result of distrust against computers substituting human reasoning and decision making.