

Borsa di Studio finanziata da fondi PNRR ex DM 351 del 09.04.2022

Missione 4, componente 1, linea di investimento 4.1

Estensione del numero di dottorati di ricerca e dottorati innovativi per la pubblica amministrazione e il patrimonio culturale
Dottorati per la Pubblica Amministrazione

Tematica di ricerca

ITALIANO Sistemi di Machine Learning scalabili, probabilistici e spiegabili per Big Data
INGLESE Scalable, probabilistic and explainable Machine Learning for Big Data

Dottorato di Ricerca

Scienze dell'ingegneria
Ciclo di dottorato: XXXVIII

Tipologia di Impresa e mesi frequenza obbligatoria

Lepida/CUP2000
Mesi 6

Mesi frequenza obbligatoria all'estero

Mesi 6

Obiettivi generali della ricerca

Il Machine Learning (apprendimento automatico, ML per brevità) è l'insieme di tecniche automatiche di analisi dei dati con cui - dai dati - si apprendono "modelli", a fini classificatori, per fare previsioni, o per raggruppare dati e individuarne - in modo automatico - le caratteristiche più salienti. La raccolta e la disponibilità di enormi quantità di dati (big data) in formato digitale, anche in nuovi contesti e scenari, apre la via all'applicazione di tecniche di analisi dei dati e di *machine learning*, che necessitano, però, di essere scalabili all'aumentare delle quantità di dati.

Il progetto di ricerca e le attività nel percorso dottorale consentiranno di formare una figura professionale in possesso di elevate competenze in ambito machine learning e data mining, con lo sviluppo e l'applicazione di algoritmi innovativi, che siano scalabili e performanti, anche su architetture HPC, per Big Data. Nel progetto formativo, sarà posta particolare attenzione al trattamento di dati sia di elevata dimensionalità, sia di dataset con classificazioni imperfette (es. informazioni mancanti) o sbilanciati in favore di una o più classi. Per trattare anche queste tipologie di dati, si considereranno algoritmi di *machine learning* probabilistico e che trattano incertezza. Per rendere fruibile il risultato dell'applicazione dei sistemi ML, particolare attenzione sarà anche dedicata alla *explainability* e intellegibilità dei modelli ottenuti.

È di immediata percezione che figure professionali, quali quelle formate in una tale percorso dottorale, possono contribuire all'accelerazione di processi di trasformazione digitale, anche in ambito pubblica amministrazione, ove siano disponibili grandi quantità di dati.

Gli algoritmi e i sistemi di ML scalabili (e probabilistici e spiegabili) definiti e sviluppati nel percorso dottorale saranno applicati in ambiti reali, forniti da *stakeholder* rilevanti della PA. Tra ambiti vi potrà essere, ad esempio, l'applicazione a (big) dati provenienti dal comparto sanitario.

English version:

Machine Learning (ML, for short) consists of automatic data analysis techniques in order to learn models from data. Usual models learned by ML are classification models, and predictive models. Also clustering is used to group data and identify their most relevant features. Nowadays, the availability of a huge amounts of data (big data), recorded in digital format, opens the way to the application of data analysis and machine learning techniques, either in new contexts and scenarios. To be effective on big data, ML techniques and algorithms must be, however, scalable as the amount of data increases.

The PhD research project will train a professional figure with high skills in the field of machine learning and data mining, and in particular with strong skills in the development and application of innovative and scalable ML algorithms, for big data and also for HPC architectures. In the PhD project, particular attention will be paid to the treatment of both high dimensional data and datasets with imperfect classifications (e.g. missing information) or unbalanced in favor of one or more classes. To deal with these types of data, probabilistic machine learning algorithms - capable to face with uncertainty - will be considered. To make the results of the application of ML systems usable by PA staff, particular attention will be also paid to learn explainable models.

The professional figure formed by this PhD program can contribute to the acceleration of digital transformation processes, even in the public administration, where large amounts of data are available.



Finanziato
dall'Unione europea
NextGenerationEU



Università
degli Studi
di Ferrara

The scalable (and also probabilistic) ML algorithms and systems defined and developed in the doctoral course will be applied in real areas, provided by relevant stakeholders of the PA. Among areas there may be, for example, the application to (big) data from the health domain.

Impatto in relazione a uno o più dei seguenti fattori: (i) miglioramento della sostenibilità ambientale; (ii) accelerazione di processi di trasformazione digitale; (iii) promozione dell'inclusione sociale

Il progetto di ricerca e le attività previste nell'ambito del percorso dottorale consentiranno di formare una figura professionale in possesso di elevate competenze nell'ambito dell'apprendimento automatico, che possa supportare vari attori e le PA *in primis*, nell'implementazione e nell'applicazione di tecniche analitiche su grandi moli di dati, anche su architetture HPC. È di immediata percezione che figure professionali di questo tipo possono contribuire all'accelerazione di processi di trasformazione digitale, che, a loro volta, consentono di velocizzare i tempi del procedimento amministrativo, come pure di aumentare l'efficienza, l'economicità e l'efficacia dell'azione amministrativa. Quest'ultima può trarre grande vantaggio dall'analisi automatica di (grandi moli di) dati, già raccolti e memorizzati in formato digitale. L'apprendimento automatico e il *data mining*, scalabili su big data, e il trattamento di dati incompleti, sbilanciati, o con incertezza, tramite approcci di ML probabilistico, rendono disponibili al decisore amministrativo e/o politico, rappresentazioni sintetiche, spiegabili e intelleggibili, estratte dai dati, che possono coadiuvarlo nel processo di *decision making*. L'impatto dell'applicazione di tali tecnologie sarà tangibile proprio in termini di consapevolezza – da parte dell'amministratore e/o decisore politico – dei modelli sottesi dai dati, che consentiranno sia di fare previsioni future sia di adottare politiche o misure opportune nel futuro.

English version:

The PhD research project will allow to train a professional figure with high skills in the field of machine learning, who can support various actors and PAs above all, in the implementation and in the application of analytical techniques on large amounts of data, also using HPC architectures. A researcher with these skills can highly accelerate the digital transformation of PA processes, and, in turn, speed up the administrative procedures, increasing efficiency, cost-effectiveness and effectiveness of the administrative action. In fact, most of administrative actions can greatly benefit from the automatic analysis of (huge amounts of) data, already collected and stored in digital format. Machine learning and data mining techniques, made scalable on big data, and the capability of automatically analyze incomplete, unbalanced, or uncertain data, through probabilistic ML approaches, will produce synthetic, explainable and intelligible representations from data, making them available to the administrative and / or political decision maker. The impact of the application of these ML techniques over the administrator and / or political decision maker will consist in the awareness about the (probabilistic, explainable) models underlying the data and extracted automatically from them. Awareness of explainable models, learned from huge amount of data, is a fundamental step for decision makers, in order to figure out future predictions and to adopt appropriate policies in the future.