

## Syllabus Attività Formativa

<b>Anno Offerta</b>	2023
<b>Corso di Studio</b>	B33 - DATA SCIENCE
<b>Regolamento Didattico</b>	B33-22-22
<b>Percorso di Studio</b>	GEN - GENERICO
<b>Insegnamento/Modulo</b>	A423699 - STATISTICAL AND MACHINE LEARNING METHODS FOR INFORMATION TECHNOLOGIES -
<b>Attività Formativa Integrata</b>	-
<b>Partizione Studenti</b>	-
<b>Periodo Didattico</b>	S1 - Primo Semestre
<b>Sede</b>	
<b>Anno Corso</b>	2
<b>Settore</b>	SECS-S/01 - STATISTICA
<b>Tipo attività Formativa</b>	B - Caratterizzante
<b>Ambito</b>	70296 - Formazione matematico-statistica
<b>CFU</b>	6.0
<b>Ore Attività Frontali</b>	48.0
<b>AF_ID</b>	178664

<b>Tipo Testo</b>	<b>Codice Tipo Testo</b>	<b>Num. Max. Caratteri</b>	<b>Ob bl.</b>	<b>Testo in Italiano</b>	<b>Testo in Inglese</b>

<b>Lingua insegnamento</b>	LINGUA_INS	100	Sì	Inglese	English
<b>Contenuti</b>	CONTENUTI	2000	Sì	<p>I principali temi del corso consistono in metodi statistici e di apprendimento automatico per la previsione dell'evoluzione di fenomeni sulla base di osservazioni, anche di grandi dimensioni prodotta dalle più moderne tecnologie</p> <p>Il corso è incentrato su l'applicazioni di modelli statistici e le tecniche di machine learning, nonché la loro interazione, allo studio di ambiti ambientali, socio-economici, bio-medicali.</p> <p>Gli argomenti trattati durante il corso sono:  Modelli lineari generalizzati (GLM):  - Regressione logistica; Regressione logistica multinomiale;  Supporta Vector Machine; alberi decisionali;  Metodi Ensemble;  Clustering;  Analisi dei dati funzionali (FDA) per la descrizione e la classificazione delle curve.</p>	<p>The main topics of the course consist of statistical methods and machine learning for the prediction of the evolution of phenomena on the basis of observations, even large ones produced by the most modern technologies</p> <p>The course focuses on the application of statistical models and machine learning techniques, as well as their interaction, to the study of environmental, socio-economic, bio-medical fields.</p> <p>The topics covered during the course are:  Generalized Linear Models (GLM):  - Logistic regression; Multinomial logistic regression;  Supports Vector Machine; decision trees;  Ensamble methods;  Clustering;  Functional data analysis (FDA) for the description and classification of the curves.</p>
<b>Testi di riferimento</b>	TESTI_RIF		Sì	Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning, 2nd ed.;	Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning, 2nd ed.;; Springer: Berlin, Germany,

				<p>Springer: Berlin, Germany, 2008; ISBN 978-0-387-84857-0</p> <p>Grant Humphries, Dawn R. Magness, Falk Huettman (Editors) Learning for Ecology and Sustainable Natural Resource Management. Published by Springer International Publishing, January 2018. DOI: 10.1007/978-3-319-96978-7, ISBNs 978-3-31-996976-3, 978-3-31-996978-7</p> <p>Articolo di riferimento:  <a href="https://www.mdpi.com/2072-4292/10/9/1365/html?msckid=9332966bd16d11ec9a73e50fecfe9613#B54-remotesensing-10-01365">https://www.mdpi.com/2072-4292/10/9/1365/html?msckid=9332966bd16d11ec9a73e50fecfe9613#B54-remotesensing-10-01365</a></p>	<p>2008; ISBN 978-0-387-84857-0</p> <p>Reference article  <a href="https://www.mdpi.com/2072-4292/10/9/1365/html?msckid=9332966bd16d11ec9a73e50fecfe9613#B54-remotesensing-10-01365">https://www.mdpi.com/2072-4292/10/9/1365/html?msckid=9332966bd16d11ec9a73e50fecfe9613#B54-remotesensing-10-01365</a></p>
<b>Obiettivi formativi</b>	OBIETT_FORM	3000	Sì	<p>Gli studenti acquisiranno competenze di tipo metodologico e applicativo e saranno in grado di applicare le conoscenze acquisite in studi su dati ambientali, ad esempio nel monitoraggio dell'inquinamento dell'aria o delle acque, nel telerilevamento per l'individuazione delle diverse tipologie di aree (urbane, agricole o industriali), nella gestione di dati di flusso, ad esempio, consumi energetici per la previsione della domanda. Inoltre, potranno anche applicare le proprie conoscenze su dati biomedici, ad esempio le</p>	<p>Students will acquire methodological and applicative skills and will be able to apply the knowledge acquired in studies on environmental data, for example in monitoring air or water pollution, in remote sensing for the identification of different types of areas ( urban, agricultural or industrial), in the management of flow data, for example, energy consumption for forecasting demand. In addition, they will also be able to apply their knowledge on biomedical data, for example functional data</p>

				tecniche di analisi di dati funzionali sull'analisi di ECG per una classificazione dei pazienti se affetti da patologia cardiaca o sani. L'obiettivo è quindi di fornire una preparazione coerente con gli sbocchi occupazionali di un data scientist in enti o imprese che operano in ambito ambientale o anche come analisti di dati biomedici in centri di cura (ospedali, laboratori bio-medicali) o in centri studio e di ricerca.	analysis techniques on ECG analysis for classifying patients with heart disease or healthy. The objective is therefore to provide a preparation consistent with the employment opportunities of a data scientist in organizations or companies operating in the environmental field or even as biomedical data analysts in treatment centers (hospitals, bio-medical laboratories) or in study centers. and research.
<b>Prerequisiti</b>	PREREQ	2000	Sì	Conoscenze di statistica di base e inferenziale e di tecniche di Data Mining. Conoscenze di base di software Python e R.	Knowledge of basic and inferential statistics and Data Mining techniques. Basic knowledge of Python and R software.
<b>Metodi didattici</b>	METODI_DID	2000	Sì	Il corso è organizzato in lezioni frontali per la presentazione delle tecniche di analisi statistica dei dati e di machine learning. Particolare attenzione è rivolta agli aspetti applicativi specialmente in ambito ambientale. L'attività di laboratorio costituisce poi un'occasione fondamentale per approfondire la conoscenza e l'utilizzo delle procedure software (Python e R) e per implementare alcuni algoritmi per analizzare dati reali o di benchmark	The course is organized in lectures for the presentation of statistical data analysis and machine learning techniques. Particular attention is paid to the application aspects, especially in the environmental field. The laboratory activity also constitutes a fundamental opportunity to deepen the knowledge and use of software procedures (Python and R) and to implement some algorithms for the analysis of real data or benchmarks.

<b>Altre informazioni</b>	ALTRO	2000	Sì		
<b>Modalità di verifica dell'apprendimento</b>	MOD_VER_AP PR	3000	Sì	<p>La verifica dell'apprendimento avviene attraverso una prova intercorso, una tesina che comprende aspetti metodologici e applicazioni delle tecniche studiate durante il corso su dati ambientali, di telerilevamento o bio-medici.</p> <p>La prova finale consiste in un colloquio su aspetti teorici delle tecniche trattate e nella presentazione della tesina; si tiene conto nella valutazione finale (espressa in trentesimi) anche del risultato della prova intercorso.</p>	<p>Learning is verified through a middle term test, a report that includes methodological aspects and applications of the techniques studied during the course on environmental, remote sensing or bio-medical data.</p> <p>The final exam consists in an interview on theoretical aspects of the techniques covered and in the presentation of the report; the final evaluation (expressed out of thirty) also takes into account the result of the middle term test.</p>
<b>Programma esteso</b>	PROGR_EST		Sì	<p>Il corso intende fornire conoscenze di metodi statistici e di apprendimento automatico per la previsione dell'evoluzione di fenomeni sulla base di osservazioni. L'enorme mole di dati prodotta dalle più moderne tecnologie ha permesso di estendere i modelli statistici e le tecniche di machine learning, nonché la loro interazione, allo studio di ambiti ambientali, socio-economici, bio-medicali.</p> <p>Il corso è incentrato su metodi statistici e metodi di machine learning per applicazioni di analisi dei dati di telerilevamento e</p>	<p>Statistics and machine learning methods concern the design of algorithms that can predict the evolution of phenomena based on observations. The huge amount of data produced by the most modern technologies has allowed to extend statistical models and machine learning techniques, as well as their interaction, to the study of environmental, socio-economics, bio-medical fields.</p> <p>The course focuses on Statistics and Machine learning methods for remote sensing data analysis applications.</p>

			<p>provenienti da sensori.</p> <p>Gli argomenti che verranno trattati durante il corso sono:</p> <p>Modelli lineari generalizzati (GLM):</p> <ul style="list-style-type: none"> <li>- Regressione logistica - quando la variabile di risposta è categoriale a due livelli (vegetazione/non vegetazione, alto/basso, presente/assente);</li> <li>- Regressione logistica multinomiale - quando la variabile di risposta ha più di due livelli (alberi/erba/terreno nudo/raccolto/acqua, alto/medio/basso)</li> </ul> <p>Supporta Vector Machine e alberi decisionali per identificare il cambiamento della copertura del suolo nelle immagini satellitari</p> <p>Telerilevamento per applicazioni ambientali:</p> <p>Metodi Ensemble:</p> <ul style="list-style-type: none"> <li>- combinare il classificatore in modo seriale</li> <li>- sistemi di classificazione multipli paralleli</li> </ul> <p>Clustering</p> <ul style="list-style-type: none"> <li>- Modello misto: modello misto gaussiano per stimare e mappare la copertura del suolo urbano</li> <li>- il metodo di clustering K-mean più utilizzato: per la riduzione della dimensione (per raggruppare i pixel dei dati rilevati) e l'aggregazione dei dati: i cluster risultanti possono essere utilizzati per ulteriori classificazioni, regressioni o altre analisi.</li> </ul>	<p>The main topics are:</p> <p>Generalised linear models (GLM):</p> <ul style="list-style-type: none"> <li>- Logistic regression - when the response variable is categorical with two levels (vegetation/not vegetation, high/low, present/absent);</li> <li>- Multinomial logistic regression - when the response variable has more than two levels (trees/grass/bare ground/crop/water, high/medium/low)</li> </ul> <p>Support Vector Machine, and decision trees to identify land cover change in satellite imagery</p> <p>Remote sensing for environmental applications:</p> <p>Ensamble methods:</p> <ul style="list-style-type: none"> <li>- combining classifier in serial fashion</li> <li>- parallel multiple classifier systems</li> </ul> <p>Clustering</p> <ul style="list-style-type: none"> <li>- Mixed model: Gaussian mixture model to estimate and map urban land cover</li> <li>- the most used K-means clustering method: for dimension reduction (to clustering pixels of sensed data) and data aggregation: resultant clusters can be used for further classification, regression or other analysis.</li> </ul> <p>Functional Data Analysis (FDA) for describing</p>
--	--	--	--	---

			Analisi dei dati funzionali (FDA) per la descrizione e la classificazione delle curve, ad esempio dei dati ECG	and classifying curves, for example of EKG data
--	--	--	--	---