



Principali informazioni sull'insegnamento

Denominazione dell'insegnamento	Data Mining	
Corso di studio	Informatica e Tecnologie per la Produzione del Software	
Anno Accademico	2022/23	
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 CFU	
Settore Scientifico Disciplinare	ING-INF/05 - Sistemi di Elaborazione delle Informazioni	
Lingua di erogazione	Italiano	
Anno di corso	Terzo	
Periodo di erogazione	2 ^a semestre, le date esatte sono riportate nel manifesto/regolamento	
Obbligo di frequenza	La frequenza è fortemente raccomandata	
Sito web del corso di studio	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/informatica-tps-270/laurea-triennale-in-informatica-e-tecnologie-per-la-produzione-del-software-d.m.-270	

Docente/i	
Nome e cognome	Paolo Mignone
Indirizzo mail	paolo.mignone@uniba.it
Telefono	+39 080 5442283
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n.571, 5 ^a piano.
Sede virtuale	Piattaforma ADA - https://elearning.di.uniba.it/
Sito web del docente	http://www.di.uniba.it/~mignone/
Ricevimento (giorni, orari e modalità, es. su appuntamento)	Venerdì 15:00 17:00 - Il ricevimento degli studenti può essere concordato via email



Syllabus			
Obiettivi formativi		<p>L'insegnamento si propone di introdurre:</p> <ul style="list-style-type: none"> • una metodologia standard per la scoperta di nuova conoscenza dai dati; • gli algoritmi di data mining più noti in letteratura; • la scelta di algoritmi di data mining per specifici compiti; • l'interpretazione dei risultati di un algoritmo di data mining <p>Lo studente acquisirà una metodologia per la scoperta di nuova conoscenza dai dati, sarà in grado di condurre le varie fasi del processo di scoperta di nuova conoscenza dai dati</p>	
Prerequisiti		<p>Le seguenti conoscenze preliminari facilitano ed accelerano la comprensione degli argomenti dell'insegnamento.</p> <p>Dall'insegnamento di Progettazione di Basi di Dati: la progettazione concettuale e logica; il linguaggio SQL.</p> <p>Dall'insegnamento di Matematica discreta: il principio di induzione, il concetto di vettori e matrici.</p> <p>Dall'insegnamento di Analisi Matematica: il concetto di funzione, variabile dipendente e indipendente</p> <p>Dall'insegnamento di Statistica per l'Ingegneria del Software: Probabilità elementare. Variabili aleatorie continue e discrete. Media, varianza e covarianza.</p>	
Contenuti di insegnamento (Programma)		<ol style="list-style-type: none"> 1. Scoperta di conoscenza nelle basi di dati: il processo. La scoperta di conoscenza nelle basi di dati: definizione. Il processo della scoperta di conoscenza nelle basi di dati. Il processo CRISP-DM: business understanding, data understanding, data preparation, modelling, evaluation, deployment. 2. Analisi di associazione. regola di associazione, itemset frequenti, principali esempi di pattern di associazione 3. Classificazione: algoritmi ad albero, basati su regole, probabilistici e basati sulla distanza. Valutazione dei modelli di classificazione 4. Regressione: algoritmi lineari e ad albero. Valutazione dei modelli di regressione 5. Analisi di raggruppamento (clustering). <p>Laboratorio: Weka data mining toolkit. Python scikit-learn.</p>	
Testi di riferimento		<p>Charu C. Aggarwal, Data Mining, Springer 2015</p> <p>Gli studenti che lo desiderano possono ottenere i testi in prestito dalla Biblioteca. Può convenire verificarne la disponibilità mediante il Sistema Bibliotecario di Ateneo https://opac.uniba.it/easyweb/w8018/index.php? e contattare la biblioteca per concordare il prestito.</p>	
Note ai testi di riferimento		<p>I testi di riferimento sono supportati da articoli scientifici e dispense forniti dal docente durante lo svolgimento del corso.</p>	
Organizzazione della didattica			
Ore			
Totali	Didattica frontale	Pratica (laboratorio, progetto, esercitazione, altro)	Studio individuale
150 ore	32 ore	30 ore	88 ore



CFU/ETCS			
6 CFU	4 CFU	2 CFU	

Metodi didattici	
	Le lezioni frontali saranno dedicate all'apprendimento dei modelli teorici e dei concetti di base, coadiuvati da alcuni esempi. Le ore di esercitazione saranno dedicate all'esecuzione di esercizi in classe coinvolgendo direttamente gli studenti nella risoluzione degli stessi.

Risultati di apprendimento previsti	
Conoscenza e capacità di comprensione	<ul style="list-style-type: none">○ Acquisizione di conoscenze relative agli algoritmi di data mining più noti in letteratura.○ Comprensione delle scelte di algoritmi di data mining per specifici compiti. Capacità di interpretazione dei risultati di un algoritmo di data mining.
Conoscenza e capacità di comprensione applicate	<ul style="list-style-type: none">○ Capacità di realizzazione di un semplice progetto di scoperta di conoscenza in una collezione di dati mediante:○ Utilizzo e/o sviluppo di strumenti per la selezione, pre-elaborazione e trasformazione dei dati, e per la validazione dei pattern estratti. Utilizzo di strumenti di data mining per l'estrazione di conoscenza finalizzata a scopi descrittivi in diversi contesti applicativi (aziendali e scientifici).
Competenze trasversali	<p>Autonomia di giudizio</p> <ul style="list-style-type: none">○ Gli studenti sono in grado di apprezzare l'uso di algoritmi di data mining in processi di scoperta della conoscenza.○ L'autonomia di giudizio viene acquisita attraverso lo studio e l'interpretazione critica dei testi.○ Il raggiungimento dell'adeguata autonomia è acquisito attraverso le esercitazioni, che si tengono durante il corso, e con l'esame finale di profitto. <p>Abilità comunicative</p> <ul style="list-style-type: none">○ Gli studenti sono in grado di esporre le tematiche incluse nel programma del corso mediante il lessico specifico della disciplina. <p>Capacità di apprendere in modo autonomo</p> <p>Gli studenti sono in grado di approfondire in autonomia le tematiche incluse nel programma del corso anche ricorrendo a risorse non direttamente coinvolte nella erogazione delle ore di lezione.</p>

Valutazione	
Modalità di verifica dell'apprendimento	Prova orale riguardante i contenuti introdotti durante le lezioni. Prova pratica, con applicazione della metodologia CRISP-DM a un insieme di dati per un particolare dominio applicativo.



Criteri di valutazione	<p>Conoscenza e capacità di comprensione:</p> <ul style="list-style-type: none">○ esposizione critica dei concetti appresi relativi al processo di scoperta della conoscenza e capacità di affrontare semplici esercizi di data mining. <p>Conoscenza e capacità di comprensione applicate:</p> <ul style="list-style-type: none">○ analisi di dataset con applicazione di algoritmi di data mining e comprensione dei risultati ottenuti, in un ciclo finalizzato al miglioramento delle prestazioni <p>Autonomia di giudizio:</p> <ul style="list-style-type: none">○ Capacità di svolgere semplici esercizi assegnati durante il corso delle lezioni <p>Abilità comunicative:</p> <ul style="list-style-type: none">○ Uso del lessico specifico della disciplina informatica <p>Capacità di apprendere:</p> <p>sviluppo di argomenti su data mining non direttamente trattati nel corso ma assegnati dal docente</p>
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	Il voto finale sarà attribuito sulla base della valutazione comparativa di tutti i criteri di valutazione sopra citati
Altro	<p>Si suggerisce agli studenti di affidarsi esclusivamente alle informazioni/comunicazioni fornite sui siti ufficiali del Dipartimento di Informatica, ovvero sui gruppi social solo se costituiti e amministrati esclusivamente dai docenti dei relativi insegnamenti:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea● https://www.uniba.it/it/ricerca/dipartimenti/informatica● https://elearning.di.uniba.it/ <p>I programmi degli insegnamenti sono disponibili qui:</p> <ul style="list-style-type: none">● https://elearning.uniba.it/course/index.php?categoryid=104 <p>Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei Regolamenti didattici e manifesti degli studi disponibili nel sito:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea <p>Si suggerisce agli studenti di diffidare delle informazioni e dei materiali circolanti su siti o gruppi social non ufficiali, poiché spesso sono risultati non affidabili, non corretti o incompleti. Per ogni dubbio, chiedere un incontro al docente secondo le modalità previste per il ricevimento.</p> <hr/> <p>https://elearning.uniba.it/course/view.php?id=2372</p>



Main information on the course

Course name	Data Mining
Degree	Bachelor's Degree in Computer Science and Technologies for Software Production
Academic year	2022/23
European Credit Transfer and Accumulation System	6 CFU
Scientific Disciplinary Sector	ING-INF/05
Lingua di erogazione	Italian
Academic year	Third
Delivery period	2nd semester, exact dates are specified in the program/regulations.
Attendance requirement	It is highly recommended to attend classes
Course of study's website	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/informatica-tps-270/laurea-triennale-in-informatica-e-tecnologie-per-la-produzione-del-software-d.m.-270

Teacher(s)	
Name and Surname	Paolo Mignone
Email	paolo.mignone@uniba.it
Phone	+39 080 5442283
Office	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Room n.571, 5th floor.
E-learning platform	Platform ADA - https://elearning.uniba.it/
Teacher's homepage	http://www.di.uniba.it/~mignone/
Office hours	Friday 15:00-17:00 - Student office hours can be scheduled via email



Syllabus			
Course goals	<p>The course aims to introduce:</p> <ul style="list-style-type: none"> • a standard methodology for discovering new knowledge from data; • the most well-known data mining algorithms in literature; • the selection of data mining algorithms for specific tasks; • the interpretation of results from a data mining algorithm. <p>Students will acquire a methodology for discovering new knowledge from data and will be capable of conducting the various phases of the data knowledge discovery process</p>		
Prerequisites/requirements	<p>The following preliminary knowledge facilitates and speeds up the understanding of the course topics.</p> <p>From the Database Design course: conceptual and logical design; the SQL language.</p> <p>From Discrete Mathematics course: the principle of induction, the concept of vectors and matrices.</p> <p>From Mathematical Analysis course: the concept of function, dependent and independent variables.</p> <p>From Software Engineering Statistics course: elementary probability. Continuous and discrete random variables. Mean, variance, and covariance.</p>		
Course program	<ol style="list-style-type: none"> 1. Knowledge discovery in databases: the process. Knowledge discovery in databases: definition. The process of knowledge discovery in databases. The CRISP-DM process: business understanding, data understanding, data preparation, modeling, evaluation, deployment. 2. Association analysis. Association rule, frequent itemsets, primary examples of association patterns. 3. Classification: tree-based algorithms, rule-based, probabilistic, and distance-based. Evaluation of classification models. 4. Regression: linear and tree-based algorithms. Evaluation of regression models. 5. Clustering analysis. <p>Lab: Weka data mining toolkit. Python scikit-learn.</p>		
Books of reference	<p>Charu C. Aggarwal, Data Mining, Springer 2015</p> <p>Students who wish to can borrow the texts from the Library. It may be advisable to check their availability through the University Library System at https://opac.uniba.it/easyweb/w8018/index.php? and contact the library to arrange the loan.</p>		
Notes to the books	<p>The reference texts are supported by scientific articles and handouts provided by the instructor during the course.</p>		
Organization of the didactic activities			
Hours			
Total	Lectures	Practice sessions	Individual study
150 hours	32 hours	30 hours	88 hours
CFU/ETCS			
6 CFU	4 CFU	2 CFU	



Teaching methods	
	<p>The frontal lessons will be dedicated to learning theoretical models and basic concepts, supported by some examples. The practice hours will be dedicated to in-class exercises directly involving students in solving them.</p>

Expected learning outcomes	
Knowledge and understanding	<ol style="list-style-type: none">1. Acquisition of knowledge related to the best-known data mining algorithms in the literature.2. Understanding of the choices of data mining algorithms for specific tasks.3. Ability to interpret the results of a data mining algorithm.
Applying knowledge and understanding	<ol style="list-style-type: none">1. Ability to carry out a simple knowledge discovery project from data collection through2. Use and/or development of tools for the selection, preprocessing, and transformation of data, and for the validation of models and patterns3. Use of data mining tools for the extraction of knowledge aimed at descriptive and predictive purposes in different application contexts (business and scientific).
Other skills	<p>Making judgements</p> <ol style="list-style-type: none">1. Students will be able to appreciate the use of data mining algorithms in knowledge discovery processes.2. Judgment autonomy is acquired through the study and critical interpretation of texts.3. Attainment of adequate autonomy is gained through exercises conducted during the course and the final exam. <p>Communicative Skills</p> <ul style="list-style-type: none">• Students can present the topics included in the course program using the specific vocabulary of the discipline. <p>Self-Learning Capability</p> <ul style="list-style-type: none">• Students are capable of independently delving deeper into the topics covered in the course, even by using resources not directly involved in the hours of instruction.

Assessment	
Assessment methods	<p>Oral exam concerning the contents introduced during the lessons. Practical test, applying the CRISP-DM methodology to a set of data for a specific application domain.</p>
Evaluation criteria	<p>Knowledge and Understanding:</p> <ul style="list-style-type: none">• Critical presentation of the concepts learned related to strategic project areas in data mining. <p>Applied Knowledge and Understanding:</p>



	<ul style="list-style-type: none">• Dataset analysis using data mining algorithms and understanding the obtained results in a cycle aimed at performance improvement. <p>Judgment Autonomy:</p> <ul style="list-style-type: none">• Ability to perform simple exercises assigned during the course. <p>Communicative Skills:</p> <ul style="list-style-type: none">• Use of specific vocabulary within the computer science discipline. <p>Learning Capability:</p> <ul style="list-style-type: none">• Development of topics on data mining not directly covered in the course but assigned by the instructor.
Measurements and final grade	The final grade will be awarded based on the comparative evaluation of all the assessment criteria mentioned above.
Further information	<p>It is recommended that students rely exclusively on information and communications provided on the official websites of the Department of Computer Science or on social groups only if they are established and managed solely by the instructors of the respective courses:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea• https://www.uniba.it/it/ricerca/dipartimenti/informatica• https://elearning.uniba.it/ <p>The course syllabi are available here:</p> <ul style="list-style-type: none">• https://elearning.uniba.it/course/index.php?categoryid=104 <p>The information that all students should know is detailed in the educational regulations and course handbooks available on the website:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea <p>Students are advised to be cautious regarding information and materials circulating on unofficial websites or social groups, as they can often be unreliable, incorrect, or incomplete. For any doubts, students should arrange a meeting with the instructor following the specified office hours.</p> <hr/> <p>Link to the course on the department's ADA e-learning platform: https://elearning.uniba.it/</p> <hr/> <p>https://elearning.uniba.it/course/view.php?id=2372</p>