



Principali informazioni sull'insegnamento

Denominazione dell'insegnamento	Data Mining II	
Corso di studio	Laurea Magistrale in Data Science	
Anno Accademico	2024/25	
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 CFU	
Settore Scientifico Disciplinare	ING/INF-05	
Lingua di erogazione	Italiano	
Anno di corso	Secondo	
Periodo di erogazione	1° semestre, le date esatte sono riportate nel manifesto/regolamento	
Obbligo di frequenza	No, ma la frequenza è fortemente raccomandata	
Sito web del corso di studio	https://www.uniba.it/it/corsi/cdl-data-science/corso-di-laurea-in-data-science	

Docente/i

Nome e cognome	Vincenzo Pasquadibisceglie
Indirizzo mail	vincenzo.pasquadibisceglie@uniba.it
Telefono	+39 080 5442407
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. V piano stanza 510.
Sede virtuale	Piattaforma e-learning UNIBA - https://elearning.uniba.it/
Sito web del docente	https://kdde.di.uniba.it/people/vincenzo-pasquadibisceglie/
Ricevimento (giorni, orari e modalità, es. su appuntamento)	Martedì dalle 10 alle 12 (su appuntamento inviando una e-mail a vincenzo.pasquadibisceglie@uniba.it)

Syllabus

Obiettivi formativi	Sviluppare competenze in merito all'utilizzo di tecniche di Data Mining con diverse tipologie di dato quali per esempio, data stream, serie temporali e log di eventi.
Prerequisiti	Conoscenza di tecniche di Data Mining di base Conoscenza del linguaggio di programmazione Python Lettura e comprensione di testi in lingua inglese
Contenuti di insegnamento (Programma)	<ol style="list-style-type: none">1) Model Centric vs Data-Centric in DM (4 ore): introduzione sui due paradigmi, descrizione pipeline di data-centric.2) Mining data streams (8 ore): Introduzione ai concetti preliminari sullo stream, gestione del cambiamento nei dati derivanti dallo stream, algoritmi per l'identificazione del cambiamento nello stream dei dati, algoritmi di classificazione nella versione online, esercizi con la libreria River.3) Mining time series data (8 ore): Introduzione alle serie temporali, statistiche mobili, test di stazionarietà e autocorrelazione, il modello ARIMA.4) Process mining (12 ore): Introduzione al Process Mining, scoperta dei modelli di processo (Process Discovery), verifica della conformità dei processi (Conformance Checking), analisi delle varianti di processo, monitoraggio predittivo dei processi (Predictive Process Monitoring), esercizi con la libreria PM4PY
Testi di riferimento	<ul style="list-style-type: none">- Aggarwal, Charu C. <i>Data mining: the textbook</i>. Vol. 1. New York: springer, 2015. per le parti 1,2,3 del programma- van der Aalst, Wil MP, and Josep Carmona. <i>Process mining handbook</i>. Springer Nature, 2022. per la parte 4 del programma <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>Nel corso delle lezioni il docente illustrerà i concetti con l'ausilio di slide che sintetizzano (e talvolta integrano) i contenuti del testo di riferimento. Le slide saranno rese disponibili al termine di ogni lezione sulla piattaforma online (v. sopra 'sede virtuale').</p> <p>Sulla piattaforma online sono disponibili:</p> <ul style="list-style-type: none"> • slide di supporto utilizzate dal docente durante le lezioni; • note del docente sul progetto richiesto per l'esame; <p>Note sul contenuto del progetto saranno fornite durante le lezioni (inoltre sulla piattaforma di e-learning alla voce Progetto esame saranno fornite indicazioni utili per lo svolgimento del progetto).</p>
-------------------------------------	---

Organizzazione della didattica	
---------------------------------------	--

Ore			
Totali	Didattica frontale	Progetto	Studio individuale
150 ore	32 ore	50 ore	68 ore
CFU/ETCS			
6 CFU	4 CFU	2 CFU	

Metodi didattici	
	Didattica in aula con lezioni di teoria.

Risultati di apprendimento previsti	
--	--

Conoscenza e capacità di comprensione	Lo studente acquisirà la conoscenza delle tecniche di Data Mining per l'analisi di data stream, serie temporali e log di eventi e la capacità di identificare le tecniche di Data Mining più appropriate per risolvere problemi quali concept drift, distribuzione sbilanciata, presenza di correlazione, disponibilità di dati multi vista, dati derivanti da modelli di processo spaghetti-like.
--	--

Conoscenza e capacità di comprensione applicate	Lo studente acquisirà la capacità di utilizzare correttamente le tecniche di Data Mining in riferimento a problemi di scoperta di conoscenza da data stream, serie temporali e log di eventi fino ad arrivare a sintetizzare o adattare nuove soluzioni di Data Mining a partire da quelle esistenti.
--	---

Competenze trasversali	<p>Autonomia di giudizio</p> <p>o Lo studente acquisirà la capacità di valutare in maniera autonoma la tipologia di dato collezionato in un problema reale e come applicare tecniche di analisi e valutazione di data mining.</p> <p>Abilità comunicative</p> <p>o Lo studente migliorerà le sue conoscenze rispetto alla capacità di usare o sintetizzare tecniche e strumenti di data mining su diverse tipologie di dato. Tali conoscenze saranno valutate durante l'esame</p>
-------------------------------	---



	<p>Capacità di apprendere in modo autonomo</p> <p>o Lo studente acquisirà l'abilità di leggere articoli scientifici su approcci di data mining pubblicati in conferenze e riviste internazionali e identificare i requisiti necessari per l'analisi di dati temporali, dati derivanti da uno stream e dati prodotti da un processo aziendale.</p>
Valutazione	
Modalità di verifica dell'apprendimento	<p>Prova scritta + Progetto</p> <p>· Prova scritta in aula, 3 domande a risposta aperta su teoria e esercizi in merito ad argomenti del syllabo; tempo assegnato 90 minuti; votazione massima 33/33. La prova scritta si ritiene superata se lo studente consegue una votazione maggiore uguale di 18/33.</p> <p>· Progetto: la consegna deve avvenire almeno una settimana prima la data dell'appello segnata su Esse3 e deve essere trasmesso tramite WeTransfer all'e-mail istituzionale del docente di riferimento. Il progetto consiste in una relazione (in PDF o Word) su un'applicazione delle tecniche di data mining in linguaggio Python. L'applicazione scelta per il progetto dovrà essere concordata preventivamente con il docente di riferimento. Il progetto sarà valutato dal docente e discusso con lo studente durante l'appello. La valutazione del progetto avverrà solo dopo il superamento della prova scritta.</p> <p>Il progetto assegnato è valido solo per gli appelli erogati nell'AA 2024-25.</p> <p>Il progetto si ritiene superato se lo studente consegue una votazione maggiore uguale di 18/33 all'atto della sua discussione.</p>
Criteria di valutazione	<p>Conoscenza e capacità di comprensione:</p> <p>o Capacità di scegliere in maniera appropriata le tecniche di Data Mining adatte in base alla natura dei dati da analizzare e la tipologia di problema da affrontare.</p> <p>• Conoscenza e capacità di comprensione applicate:</p> <p>o Capacità di descrivere, con esempi pratici del mondo reale, le varie tecniche di data mining per la gestione di serie temporali, dati derivanti da uno stream e dati prodotti da processi aziendali.</p> <p>Autonomia di giudizio:</p> <p>o Abilità di valutare i risultati conseguiti applicando una pipeline di Data Mining identificando limiti e possibili miglioramenti.</p> <p>• Abilità comunicative:</p> <p>o Capacità di descrivere, attraverso gli argomenti trattati nel corso, le scelte intraprese per gestire le varie tipologie di dato e scegliere le tecniche di data mining idonee.</p> <p>• Capacità di apprendere:</p> <p>o Comprensione di articoli scientifici che descrivono applicazioni di approcci di Data Mining a problemi che coinvolgono dati di diversa natura (serie temporali, stream di dati, ecc.).</p>
Criteria di misurazione dell'apprendimento e di attribuzione del voto finale	<p>Il voto finale è determinato con la seguente formula $1/2*(votazione\ prova\ scritta) + 1/2*(votazione\ del\ progetto)$.</p>
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• https://www.uniba.it/it/ricerca/dipartimenti/informatica• https://elearning.uniba.it/



I programmi di tutti gli insegnamenti sono disponibili al seguente link:

- <https://elearning.uniba.it/>

Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei regolamenti didattici dei Corsi di Studi disponibili nel sito:

- <https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea>

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.

Gli studenti potranno unirsi al forum del corso A.A. 2024/25 iscrivendosi al corso sulla piattaforma e-learning:

<https://elearning.uniba.it/>



Main information on the course	
Course name	Data Mining II
Degree	Master's Degree in Data Science
Academic year	2024/25
European Credit Transfer and Accumulation System (ECTS), in Italian Crediti Formativi Universitari (CFU)	6 CFU (each CFU corresponds to 25 hours (h) of student's time); CFU are of type T1, T2 or T3 T1 = 8 h lecture + 17 h individual study T2 = 15 h practice + 10 h individual study T3 = 25 h individual study
SSD	ING-ING/05
Course language	Italian
Course year	Second
Course period	First Semester/Second Semester - exact dates can be found in the didactic regulations
Course attendance requirement	None, but it is highly recommended to attend classes
Website of the Degree	https://www.uniba.it/it/corsi/cdl-data-science/corso-di-laurea-in-data-science

Teacher(s)	
Name and Surname	Vincenzo Pasquadibisceglie
email	vincenzo.pasquadibisceglie@uniba.it
phone	+39 080 5442407
office	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. V piano stanza 510.
e-learning platform	Piattaforma e-learning UNIBA - https://elearning.uniba.it/
Teacher's homepage	https://kdde.di.uniba.it/people/vincenzo-pasquadibisceglie/
Office hours	Martedì dalle 10 alle 12 (su appuntamento inviando una e-mail a vincenzo.pasquadibisceglie@uniba.it)

Syllabus	
Course goals	Develop skills in using data mining techniques with different types of data, such as data streams, time series, and event logs.
Prerequisites/requirements	Basic concepts in data mining techniques Python programming language Reading and understanding text in English
Course program	1) Model-centric vs. Data-Centric in DM (4 hrs): An introduction to the two paradigms and a pipeline description of data-centricity. 2) Mining data streams (8 hours): An introduction to preliminary concepts on streams, handling data from the stream, algorithms for identifying change in the data stream, classification algorithms in the online version, and exercises with the River library. 3) Mining time series data (8 hours): An introduction to time series, moving statistics, stationarity and autocorrelation tests, and the ARIMA model. 4) Process mining (12 hours): Introduction to Process Mining, process pattern discovery (Process Discovery), process conformity checking (Conformance Checking), process variant analysis, predictive process monitoring (Predictive Process Monitoring), exercises with the PM4PY library
Books of reference	- Aggarwal, Charu C. Data mining: the textbook. Vol. 1. New York: Springer, 2015. for parts 1,2,3. - van der Aalst, Wil MP, and Josep Carmona. Process mining handbook. Springer Nature, 2022. for part 4 of the programme



	Students who wish to do so may obtain the texts on loan from the Library. You may wish to check their availability through the University Library System https://opac.uniba.it/easyweb/w8018/index.php? and contact the Library to arrange the loan.		
Notes to the books	<p>During the lectures, the lecturer illustrates concepts with slides that summarise (and sometimes supplement) the contents of the reference text. The slides are made available at the end of each lecture on the online platform (see 'virtual venue' above). Available on the online platform are:</p> <ul style="list-style-type: none"> • supporting slides used by the lecturer during the lectures; • lecturer's notes on the project required for the exam; <p>Notes on the project's content will be provided during the lectures (in addition, valuable hints for the project will be provided on the e-learning platform under the Examination project).</p>		
Organization of the didactic activities			
Hours			
Total	Lectures	Project work	Individual study
150 hours	32 hours	50 hours	68 hours
CFU/ETCS			
6 CFU	4 CFU	2 CFU	

Teaching methods	
	Classroom teaching with theory lessons.

Expected learning outcomes	
Knowledge and understanding	The student will acquire knowledge of Data Mining techniques for analysing data streams, time series, and event logs and the ability to identify the most appropriate techniques to solve problems such as concept drift, unbalanced distribution, presence of correlation, multi-view data availability, and data from spaghetti-like process models.
Applying knowledge and understanding	The student will acquire the ability to correctly use data mining techniques to solve knowledge discovery problems from data streams, time series, and event logs to the point of synthesising or adapting new data mining solutions from existing ones.
Other skills	<p><i>Making judgements</i> The student will be able to independently evaluate the type of data collected in a real-world problem and how to apply data mining analysis and evaluation techniques.</p> <p><i>Communication</i> The student will improve their knowledge of using or synthesising data mining techniques and tools on different data types. This knowledge will be assessed during the exam.</p> <p><i>Learning skills</i> The student will acquire the ability to read scientific articles on data mining approaches published in international conferences and journals and identify the requirements for analysing temporal data, data from a stream and data produced by a business process.</p>

Assessment	
Assessment methods	<p>Written test + Project</p> <ul style="list-style-type: none"> · Written test in the classroom, 3 open-ended questions on theory and exercises on syllabus topics; time allocated 90 minutes; maximum mark 33/33. The written test is deemed to have been passed if the student achieves a mark higher than or equal to 18/33. · Project: delivery must take place at least one week before the date of the roll call marked on Esse3 and must be transmitted via WeTransfer to the institutional e-mail address of the teacher of reference. The project consists of a report (in PDF or Word) on an application of data mining techniques in the Python language. The application chosen for the project must be agreed in



	<p>advance with the reference lecturer. The project will be evaluated by the lecturer and discussed with the student during the call.</p> <p>The project will only be assessed after the written test has been passed.</p> <p>The assigned project is only valid for appeals held in the AA 2024-25.</p> <p>The project is deemed to have been passed if the student obtains a mark greater than or equal to 18/33 when it is discussed.</p>
Evaluation criteria	<p>Knowledge and comprehension skills:</p> <ul style="list-style-type: none">- Ability to appropriately choose suitable Data Mining techniques based on the nature of the data to be analysed and the type of problem to be addressed.• Applied knowledge and understanding:- Ability to describe, with real-world practical examples, the various data mining techniques for handling time series, data derived from a stream and data produced by business processes. <p>Autonomy of judgement:</p> <ul style="list-style-type: none">- Ability to evaluate the results achieved by applying a data mining pipeline by identifying limitations and possible improvements.• Communication skills:- Ability to describe, through the topics covered in the course, the choices made to manage the various types of data and choose suitable data mining techniques.• Learning skills:- Understanding of scientific articles describing applications of data mining approaches to problems involving data of different types (time series, data streams, etc.).
Measurements and final grade	<p>The final grade is determined using the following formula $1/2*(\text{written test mark}) + 1/2*(\text{project mark})$.</p>
Further information	<p>Students are advised to rely exclusively on the information/communication provided on the official websites of the Department of Computer Science, or on social groups only if they are set up and administered exclusively by the lecturers of the relevant subjects:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea• https://www.uniba.it/it/ricerca/dipartimenti/informatica• https://elearning.uniba.it/ <p>The syllabuses of all courses are available at the following link:</p> <ul style="list-style-type: none">• https://elearning.uniba.it/ <p>Information that all students should be aware of is written in the teaching regulations of the study courses available on the website:</p> <ul style="list-style-type: none">• https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea <p>Students are advised to be wary of information and materials circulating on unofficial websites or social groups, as they are often found to be unreliable, incorrect or incomplete. If you have any doubts, please ask the lecturer for a meeting in accordance with the reception arrangements.</p> <hr/> <p>Students can join the A.A. 2024/25 course forum by registering for the course on the e-learning platform: https://elearning.uniba.it/</p>