## Principali informazioni sull'insegnamento

| | |
|---|---|
| Denominazione dell'insegnamento | **Analisi dei Dati per la Sicurezza** |
| Corso di studio | Laurea Magistrale in Sicurezza Informatica |
| 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 | Primo |
| Periodo di erogazione | 1^ 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/sicurezza-informatica/laurea-magistrale-in-informatica |

## Docente/i



| | |
|---|---|
| Nome e cognome | Annalisa Appice |
| Indirizzo mail | annalisa.appice@uniba.it |
| Telefono | 080544 3262 |
| Sede | Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n. 512, V piano. |
| Sede virtuale | Piattaforma - https://elearning.uniba.it/ |
| Sito web del docente | http://www.di.uniba.it/~appice/ |
| Ricevimento (giorni, orari e modalità, es. su appuntamento) | Su appuntamento |



## Syllabus

| | |
|---|---|
| **Obiettivi formativi** | Il corso si propone di fornire gli strumenti algoritmici per lo sviluppo di competenze teoriche in merito al metodo scientifico di indagine, metodi, tecniche e strumenti per l'analisi dei cyber dati |
| **Prerequisiti** | Concetti di base di programmazione in Python, calcolo delle probabilità (distribuzione di probabilità, probabilità condizionata, T-student test), calcolo numerico (matrici, autovalori, autovettori)<br>Lettura e comprensione di testi in lingua inglese |
| **Contenuti di insegnamento (Programma)** | Introduzione: Sicurezza Informatica, Data Mining per Sicurezza Informatica (4 ore)<br><br>KDD e Data Mining: paradigmi tradizionali di data mining e analisi di dati (processo KDD, fondamenti dei paradigmi di apprendimento supervisionati, non supervisionati e semi-supervisionati; compiti e metodi fondamentali, metodi di selezione di feature e sampling, tecniche di valutazione (12 ore)<br>Metodi di data mining per la classificazione, clustering, anomaly detection (16 ore)<br><br>Laboratorio: Esercizi sull'applicazione di tecniche di KDD e Data Mining. Programmazione in Python usando librerie di sklearn (15 ore) |
| **Testi di riferimento** | **Teoria:**<br>Max Bramer, Principles of Data Mining. Terza edizione, Springer, 2016 **(pagine 1-187, 209-236, 311-328)**<br><br>Christopher M. Bishop, Pattern Recognition and Machine Learning by, 2018 **(capitoli 11,12)**<br><br>**Laboratorio:**<br><br>Raschka, Sebastian - Mirjalili, Vahid Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 / Sebastian Raschka, Vahid Mirjalili. - 3. ed. - Birmingham ; Mumbai : Packt, 2019. - xxi, 741 p. : ill. ; 24 cm<br><br>**Testi per approfondimenti**<br><br>**Aggarwal, Charu C. Neural Networks and Deep Learning A Textbook, 2018 https://www.springer.com/gp/book/9783319944623**<br><br>Sarah Guido, Andreas C. Müller. Introduction to Machine Learning with Python. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/<br><br>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. |
| **Note ai testi di riferimento** | Nel corso delle lezioni il docente illustrerà i concetti con l'ausilio di slide che sintetizzano (e talvolta integrando) i contenuti del testo di riferimento e riportano esercizi svolti in aula. Le slide saranno rese disponibili al termine di ogni lezione sulla piattaforma ADA del dipartimento (v. sopra 'sede virtuale').<br><br>Sulla piattaforma ADA sono disponibili:<br>● slide di supporto utilizzate dal docente durante le lezioni;<br>● esercizi con soluzioni;<br>● alcune tracce di prove scritte di esami precedenti. |

| | |
|---|---|
| | La traccia del progetto sarà fornita durante le lezioni di laboratorio che inizieranno approssimativamente a fine ottobre 2024 (la traccia sarà pubblicata su ADA). |

## Organizzazione della didattica

**Ore**

| Totali | Didattica frontale | Pratica (laboratorio, progetto, esercitazione, altro) | Studio individuale |
|---|---|---|---|
| 47 ore | 32 ore | 15 ore | 78 ore |

**CFU/ETCS**

| | | | |
|---|---|---|---|
| 6 CFU | 4 CFU | 4 CFU | |

## Metodi didattici

Didattica in aula con lezioni di teoria, esercitazioni guidate durante le quali gli studenti interagiranno con il docente per lo svolgimento degli esercizi, attività di laboratorio per la programmazione e svolgimento di progetto (eventualmente in gruppo). I gruppi possono includere un numero di componenti che varia da 1 a 3. L'organizzazione in gruppi è decisa liberamente dagli studenti, ma va comunicata al docente al momento della consegna del progetto e non può essere successivamente modificata.
L'attività di esercitazione, laboratorio, e progetto sono finalizzate ad affiancare pratica e teoria per lo sviluppo di software e soluzione di problemi con l'uso del computer

## Risultati di apprendimento previsti

| | |
|---|---|
| **Conoscenza e capacità di comprensione** | La classe acquisirà l'abilità di elaborare cyber dati con tecniche di analisi dei dati con l'obiettivo di scoprire cyber minacce. Si focalizzerà l'attenzione su tecniche di knowledge discovery (KDD) e algoritmi di data mining per compiti supervisionati e non-supervisionati. |
| **Conoscenza e capacità di comprensione applicate** | La classe sarà in grado di utilizzare la conoscenza acquisita per definire e applicare una pipeline KDD per risolvere un problema di cybersecurity e interpretare i risultati conseguiti. |
| **Competenze trasversali** | *Autonomia di giudizio* |

| | |
|---|---|
| | ● *La classe acquisterà l'abilità di analizzare i cyber dati con tecniche di KDD e data mining con l'obiettivo di scoprire cyber minacce. Durante l'esame lo studente dovrà discutere gli approcci di KDD e data mining oggetto del programma e risolvere un problema di cybersecurity applicando una pipeline KDD.*<br><br>*Abilità comunicative*<br><br>● *La classe migliorerà le sue conoscenze rispetto alla capacità di usare strumenti di KDD e data mining in problemi di cybersecurity. Tali conoscenze saranno valutate durante l'esame*<br><br>● *Capacità di apprendere in modo autonomo*<br><br>● *La classe acquisirà l'abilità di leggere articoli scientifici su approcci di KDD e data mining pubblicati in conferenze e riviste internazionali, come anche progettare nuove pipeline KDD per sistemi di cyber difesa.* |

| **Valutazione** | |
|---|---|
| **Modalità di verifica dell'apprendimento** | Prova scritta + Progetto<br><br>· Prova scritta in aula, 3 domande aperte su teoria ed esercizi in merito ad argomenti del sillabo; 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. È consentito l'uso della calcolatrice tradizionale durante la prova (no pc, smartphone, tablet).<br><br>· Progetto: la consegna deve avvenire con weTransfer indirizzato alla email istituzionale del docente di riferimento del corso entro 24 ore prima la discussione del progetto e deve includere il codice prodotto in Python e la presentazione power point dell'attività di progetto svolta. Durante la discussione del progetto lo studente ha 10 minuti per presentare il progetto con le slide power point e dopo è chiamato a rispondere a domanda del docente in merito al modo in cui è stato scritto il codice corrispondente allo sviluppo del progetto.<br><br>Il codice del progetto può essere sviluppato in gruppo (massimo 3 componenti), ma la presentazione power-point del progetto è distinta per ogni studente. La discussione del progetto è individuale.<br><br>Inoltre, il progetto prevede un'attività assegnata a ciascuno studente (non svolta in gruppo). Lo studente dovrà contattare il docente per avere l'assegnazione dell'attività personale. |

| | |
|---|---|
| | Il progetto può essere discusso solo dopo il superamento della prova scritta.<br><br>Il progetto assegnato è valido solo per gli appelli erogati nell'AA 2024-25.<br><br>La traccia del progetto è assegnata durante le lezioni di laboratorio (e contestualmente pubblicata su ADA).<br><br>Il progetto si ritiene superato se lo studente consegue una votazione maggiore uguale di 18/33 all'atto della sua discussione. |
| Criteri di valutazione | Conoscenza e capacità di comprensione:<br><br>o Capacità di scegliere in maniera appropriata le tecniche di KDD e l'algoritmo di data mining adatto per un dato problema di cybersecurity.<br><br>Conoscenza e capacità di comprensione applicate:<br><br>o Capacità di spiegare le differenze tra diversi algoritmi di data mining e tecniche di KDD.<br><br>Autonomia di giudizio:<br><br>o Abilità di valutare i risultati conseguiti applicando una pipeline di KDD identificando limiti e possibili miglioramenti.<br><br>Abilità comunicative:<br><br>o Capacità di motivare la definizione di una specifica pipeline di KDD nella risoluzione di un problema di cybersecurity e illustrare i risultati conseguiti applicando la pipeline formulata.<br><br>Capacità di apprendere:<br><br>Comprensione di articoli scientifici che descrivono applicazioni di approcci di KDD e data mining a problemi di cybersecurity. |
| Criteri di misurazione dell'apprendimento e di attribuzione del voto finale | Il voto finale è determinato con la seguente formula 4/6*(votazione prova scritta) + 2/6*(votazione a seguito di discussione del progetto), arrotondato per eccesso. |
| **Altro** | 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:<br><br>● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea<br><br>● https://www.uniba.it/it/ricerca/dipartimenti/informatica |

- https://elearning.di.uniba.it/

I programmi degli insegnamenti sono disponibili qui:

- https://programmi.di.uniba.it/

Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei Regolamenti didattici e manifesti degli studi disponibili nel sito:

- https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/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 del dipartimento ADA: https://elearning.di.uniba.it/

## Main information on the course

| | | |
|---|---|---|
| Course name | **Data Analysis for Security** | |
| Master degree | Master degree in Computer Security | |
| Academic Year | 2024/25 | |
| University credits (CFU) / European Credit Transfer and Accumulation System (ECTS) | | 6 CFU |
| SSD | ING-INF 05 | |
| Language | English | |
| Year | First | |
| Time | 1^ semester, the exact dates are shown in the poster/regulations | |
| Attendance | Attendance is strongly recommended | |
| Web site | https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/sicurezza-informatica/laurea-magistrale-in-informatica | |

## Teacher



| | |
|---|---|
| Name and surname | Annalisa Appice |
| Email | annalisa.appice@uniba.it |
| Phone | 080544 3262 |
| Location | Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n. 512, V piano. |
| Virtual location | https://elearning.uniba.it/ |
| Teacher Web site | http://www.di.uniba.it/~appice/ |
| Meetings | On appointment |

| Syllabus | |
|---|---|
| **Training objectives** | The course aims to provide algorithmic tools for the development of theoretical skills regarding the scientific method of investigation, methods, techniques and tools for the analysis of cyber data |
| **Prerequisites** | Basic programming concepts in Python, probability calculus (probability distribution, conditional probability, T-student test), numerical computation (matrices, eigenvalues, eigenvectors)<br>Reading and comprehension of texts in English |
| **Programme** | Introduction: IT Security, Data Mining for IT Security (4 hours)<br><br>KDD and Data Mining: traditional data mining and data analysis paradigms (KDD process, foundations of supervised, unsupervised and semi-supervised learning paradigms; fundamental tasks and methods, feature selection and sampling methods, evaluation techniques (12 hours)<br>Data mining methods for classification, clustering, anomaly detection (16 hours)<br><br>Laboratory: Exercises on the application of KDD and Data Mining techniques. Programming in Python using sklearn libraries (15 hours) |
| **Books** | Theory:<br>Max Bramer, Principles of Data Mining. Third edition, Springer, 2016 (pages 1-187, 209-236, 311-328)<br><br>Christopher M. Bishop, Pattern Recognition and Machine Learning by, 2018 (chapters 11,12)<br>Laboratory:<br>Raschka, Sebastian - Mirjalili, Vahid Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 / Sebastian Raschka, Vahid Mirjalili. - 3rd ed. - Birmingham ; Mumbai : Packt, 2019. - xxi, 741 p. : ill. ; 24cm<br><br>Books for further information<br>Aggarwal, Charu C. Neural Networks and Deep Learning A Textbook, 2018 https://www.springer.com/gp/book/9783319944623<br>Sarah Guido, Andreas C. Müller. Introduction to Machine Learning with Python. https://www.oreilly.com/library/view/introduction-to-machine/9781449369880/<br><br>Students who wish can borrow texts from the Library. Is it convenient to check availability via the University Library System https://opac.uniba.it/easyweb/w8018/index.php? and contact the library to arrange the loan. |
| **Notes to books** | During the lessons, the teacher will illustrate the concepts with the help of slides that summarize (and sometimes integrate) the contents of the reference text and report exercises carried out in the classroom. The slides will be made available at the end of each lesson on the department's ADA platform (see 'virtual venue' above).<br><br>The following are available on the ADA platform: |

| | |
|---|---|
| | ● support slides used by the teacher during lessons;<br>● exercises with solutions;<br>● some traces of written tests from previous exams.<br><br>The project outline will be provided during the laboratory lessons starting approximately at the end of October 2024 (the outline will be published on ADA). |

## Organization of lectures

| **Hours** | | | |
|---|---|---|---|
| Total | Frontal teaching | Practice (laboratory, project, exercise, other) | Individual study |
| 47 hours | 32 hours | 15 hours | 78 hours |
| **CFU/ETCS** | | | |
| 6 CFU | 4 CFU | 4 CFU | |

## Teaching Methods

Classroom teaching with theory lessons, guided exercises during which students will interact with the teacher to carry out the exercises, laboratory activities for planning and carrying out the project (possibly in a group). The groups can include a number of members ranging from 1 to 3. The organization into groups is freely decided by the students, but must be communicated to the teacher at the time of delivery of the project and cannot be subsequently modified.
The exercise, laboratory and project activities are aimed at combining practice and theory for the development of software and problem solving with the use of computers

## Expected Learning Results

| | |
|---|---|
| **Knowledge and Understanding** | The class will acquire the ability to process cyber data with data analysis techniques with the aim of discovering cyber threats. We will focus on knowledge discovery (KDD) techniques and data mining algorithms for supervised and unsupervised tasks. |
| **Applied Knowledge and Understanding** | The class will be able to use the knowledge acquired to define and apply a KDD pipeline to solve a cybersecurity problem and interpret the results achieved. |
| **Transversal skills** | Independent judgment<br>● The class will acquire the ability to analyze cyber data with KDD and data mining techniques with the aim of discovering cyber threats. During the exam the |

| | |
|---|---|
| | student will have to discuss the KDD and data mining approaches covered by the program and solve a cybersecurity problem by applying a KDD pipeline.<br>Communication skills<br>● The class will improve its knowledge regarding the ability to use KDD and data mining tools in cybersecurity problems. This knowledge will be assessed during the exam<br>Ability to learn independently<br>● The class will acquire the ability to read scientific articles on KDD and data mining approaches published in international conferences and journals, as well as design new KDD pipelines for cyber defense systems. |

| **Evaluation** | |
|---|---|
| **Learning assessment methods** | Written test + Project<br>● Written test in the classroom, 3 open questions on theory and exercises regarding the topics of the syllabus; allotted time 90 minutes; maximum score 33/33. The written test is considered passed if the student achieves a grade greater than 18/33. The use of a traditional calculator is permitted during the test (no PCs, smartphones, tablets).<br>● Project: delivery must take place with weTransfer addressed to the institutional email of the course reference teacher within 24 hours before the discussion of the project and must include the code produced in Python and the power point presentation of the project activity carried out. During the discussion of the project, the student has 10 minutes to present the project with Power Point slides and is then asked to answer the teacher's question regarding the way in which the code corresponding to the development of the project was written.<br>The project code can be developed in a group (maximum 3 members), but the power-point presentation of the project is separate for each student. The discussion of the project is individual.<br>**In addition, the project includes an activity that is assigned to each student (not performed in group). The student must contact the teacher to have the personnel activity assignment.**<br>The project can only be discussed after passing the written test.<br>The project outline is assigned during the laboratory lessons (and simultaneously published on ADA).<br>The project is considered passed if the student achieves a grade greater than 18/33 at the time of its discussion.<br><br>The assigned project is valid only for the exam sessions held in the 2024-25 academic year. |
| **Evaluation Criteria** | Knowledge and understanding:<br>o Ability to appropriately choose the KDD techniques and the data mining algorithm suitable for a given cybersecurity problem.<br>Applied knowledge and understanding:<br>o Ability to explain the differences between different data mining algorithms and KDD techniques.<br>Independent judgment:<br>o Ability to evaluate the results achieved by applying a KDD pipeline, identifying limitations and possible improvements.<br>Communication skills:<br>o Ability to motivate the definition of a specific KDD pipeline in solving a cybersecurity problem and illustrate the results achieved by applying the formulated pipeline.<br>Ability to learn: |

| | |
|---|---|
| | Understanding of scientific articles describing applications of KDD and data mining approaches to cybersecurity problems. |
| Evaluation Measurement Criteria and Final Grade | The final grade is determined with the following formula 4/6*(written test grade) + 2/6*(grade following discussion of the project), rounded up. |
| **Additional information** | Students are advised to rely exclusively on the information/communications provided on the official websites of the Department of Computer Science, or on social groups only if established and administered exclusively by the teachers of the relevant courses:<br><br>● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea<br>● https://www.uniba.it/it/ricerca/dipartimenti/informatica<br>● https://elearning.di.uniba.it/<br><br>The teaching programs are available here:<br>● https://programmi.di.uniba.it/<br><br>The information that all students should know is written in the Teaching Regulations and study posters available on the site:<br>● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea<br><br>Students are advised to be wary of information and materials circulating on unofficial sites or social groups, as they are often found to be unreliable, incorrect or incomplete. If you have any doubts, ask the teacher for a meeting according to the reception procedures.<br><br>_____<br><br>Students will be able to join the A.A. course forum. 2024/25 by enrolling in the course on the e-learning platform of the ADA department: https://elearning.di.uniba.it/ |