Principali informazioni sull'insegnamento	
Titolo insegnamento	Analisi di Dati per la Sicurezza
Corso di studio	Magistrale in Sicurezza Informatica
Crediti formativi	4+1+1
Denominazione inglese	Data Analytics for Security
Obbligo di frequenza	No
Lingua di erogazione	Italiano

Docente responsabile	Nome Cognome	Indirizzo Mail
	Annalisa Appice	annalisa.appice@uniba.it

Dettaglio credi formativi	Ambito disciplinare	SSD	Crediti
	Informatico	ING INF 05	4+1+1

Modalità di erogazione	
Periodo di erogazione	Semestre 1
Anno di corso	1
Modalità di erogazione	Lezioni frontali 47 ore (32 Teoria 15 Esercitazione e/o Laboratorio)

Organizzazione della didattica	
Ore totali	125 ore
Ore di corso	47
Ore di studio individuale	78

Calendario	
Inizio attività didattiche	26 Settembre 2017
Fine attività didattiche	13 Gennaio 2018

Syllabus	
Prerequisiti	*Basi di Dati, Algoritmi, Logica, Statistica, Teoria dell'Informazione.*
Risultati di apprendimento previsti (declinare rispetto ai Descrittori di Dublino) (si raccomanda che siano coerenti con i risultati di apprendimento del CdS, compreso i risultati di apprendimento trasversali)	• *Conoscenza e capacità di comprensione* Lo studente dovrà acquisire una conoscenza di base della tecniche di data mining e applicazione delle stesse in analisi di dati collezionati in applicazioni di sicurezza informatica. • *Conoscenza e capacità di comprensione applicate* Attraverso l'introduzione e la applicazione di diverse tecniche di data mining lo studente acquisirà capacità applicative sullo utilizzo di strumenti per analisi di dati per la estrazione e validazione di pattern di conoscenza estratti in diversi contesti applicativi nella sicurezza informatica. • *Autonomia di giudizio*

	Lo studente dovrà dimostrare di aver acquisito una elevata autonomia di giudizio nella definizione e valutazione di pipeline di servizi di data mining per la l'analisi di dati rinvenuti da applicazioni di sicurezza informatica • *Abilità comunicative* Lo studente sarà in grado di relazionare in maniera appropriate in riferimento alle tecniche di data mining e alla progettazione e applicazione di una pipeline di data mining per l'analisi di dati nell'ambito di applicazioni di sicurezza informatica. • *Capacità di apprendere* Lo studente dovrà mostrare di aver sviluppato capacità di apprendere e di orientarsi agilmente nelle problematiche che si presentano durante la analisi di dati rinvenuti in diverse tipologie di applicazioni di Sicurezza Informatica
Contenuti di insegnamento	Introduzione: Sicurezza Informatica, Data Mining per Sicurezza Informatica 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, tecniche di valutazione, metodi di data mining per la classificazione, clustering, scoperta di associazioni, anomaly detection) Data mining in Malware Detection Data mining in Intrusion Detection (misuse e anomaly detection) Data mining in Phishing Detection Data mining in Spam Detection (mail spam detection e review spam detection) Data mining in Credit Card Fraud detection, Authentication

Programma	
Testi di riferimento	Maimon, Oded, Rokach, Lior (Eds.) Data Mining and Knowledge Discovery Handbook, Springer 2010 Ian H. Witten, Eibe Frank. Data Mining: Practical Machine Learning Tools and Techcniques, Elsevier, Second Edition Varun Chandola, Arindam Banerjee, Vipin Kumar. Anomaly detection: A Survey. Technical report. Yanfang Ye, Tao Li, Donald Adjeroh, S. Sitharama Iyengar. A Survey on Malware Detection Using Data Mining Techniques.

	ACM Computing Surveys, Vol. 50, No. 3, Article 41, June 2017. Gandotra, E., et al. (2014) Malware Analysis and Classification: A Survey. Journal of Information Security, 5, 56-64. Anna L. Buczak, and Erhan Guven. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection, IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 18, NO. 2, SECOND QUARTER 2016, pp 1153-1176 Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Phishing Detection: A Literature Survey. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 15, NO. 4,pp. 2091-2121, 2013 Jyoti Vaibhav Jadhav, Prof. Krishna Kumar Tripathi. A Survey of Phishing Website Detection Techniques. International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2017 Enrico Blanzieri and Anton Bryl. 2008. A survey of learning-based techniques of email spam filtering. *Artif. Intell. Rev.* 29, 1 (March 2008), 63-92. DOI: https://doi.org/10.1007/s10462-009-9109-6 Anjali Sharm, Manisha, Rekha Jain. A survey on spam detection techniques. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 12, December 2014 Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter and Hamzah Al Najada. Survey of review spam Detection using machine learning techniques. Crawford et al. Journal of Big Data (2015) 2:23 DOI 10.1186/s40537-015-0029-9 S.Vimala1, K.C.Sharmili. Survey Paper for Credit Card Fraud Detection Using Data Mining Techniques. International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 6, Special Issue 11, September 2017 Daniel Barbara and Sushil Jajodia. Applications of Data mining in Computer Security. Kluwer Academoc Publisher, 2002 Sumeet Dua and Xian Du Data Mining and Machine Learning in CyberSecurity, CRC Press 2011
Note ai testi di riferimento	I testi di riferimento saranno integrati con slide e materiale didattico messo a disposizione dal docente sulla piattaforma ADA
Metodi didattici	Lezioni frontali ed esercitazioni pratiche
Metodi di valutazione (indicare almeno la tipologia scritto, orale, altro)	Prova orale e discussione di un caso di studio relativo alla applicazione di tecniche di data mining nella analisi di dati per la sicurezza informatica.

Criteri di valutazione (per ogni risultato di apprendimento atteso su indicato, descrivere cosa ci si aspetta lo studente conosca o sia in grado di fare e a quale livello al fine di dimostrare che un risultato di apprendimento è stato raggiunto e a quale livello)	In considerazione della natura teorico-pratica del corso, la verifica dell'apprendimento avverrà durante la prova orale. *Conoscenza e capacità di comprensione* Durante la prova orale lo studente dovrà rispondere a quesiti che verificheranno la sua padronanza dei modelli e tecniche di data mining e la loro applicazione in applicazioni di sicurezza informatica *Conoscenza e capacità di comprensione applicate* Durante lo sviluppo del caso di studio dovrà individuare la pipeline di data mining più appropriata alla risoluzione di un problema formulato dal docente. *Autonomia di giudizio* Lo studente dovrà utilizzare gli strumenti di validazione per valutare la bontà, significatività a innovatività dei pattern scoperti. *Abilità comunicative* Lo studente dovrà dimostrare durante la prova orale proprietà di linguaggio e padronanza dei contenuti del corso. *Capacità di apprendere* Lo studente dovrà dimostrare la sua capacità di apprendere tramite la applicazione di tecniche di data mining in un caso di studio.
Altro	