FIȘA DISCIPLINEI

1. Date despre program

1.1 Instituţia de învăţământ superior	Universitatea Tehnică din Cluj-Napoca
1.2 Facultatea	Automatică și Calculatoare
1.3 Departamentul	Calculatoare
1.4 Domeniul de studii	Calculatoare si Tehnologia Informatiei
1.5 Ciclul de studii	Master
1.6 Programul de studii / Calificarea	Securitatea Informatiilor si Sistemelor de calcul/ Master
1.7 Forma de învăţământ	IF – învăţământ cu frecvenţă
1.8 Codul disciplinei	9.1

2. Date despre disciplină

2.1 Denumirea disciplinei				***Sisteme de date masive și securitatea calculatoarelor***	
2.2 Titularii de curs				Conf.dr.ing. Camelia LEMNARU (camelia.lemnaru@cs.utcluj.ro)	
2.3 Titularul/Titularii activităţilor de seminar/laborator/proiect				Conf.Dr.ing. Ciprian OPRIȘA (ciprian.oprisa@cs.utcluj.ro)	
2.4 Anul de studiu	I	2.5 Semestrul	2	2.6 Tipul de evaluare *(E – examen, C – colocviu, V – verificare)*	E
2.7 Regimul disciplinei	*DA – de aprofundare, DS – de sinteza, DC – complementară*				DS
	DI – Impusă, DOp – opțională, DFac – facultativă				DOp

3. Timpul total estimat

3.1 Număr de ore pe săptămână	4	din care:	Curs	2	Seminar		Laborator	2	Proiect	
3.2 Număr de ore pe semestru	56	din care:	Curs	28	Seminar		Laborator	28	Proiect	
3.3 Distribuţia fondului de timp (ore pe semestru) pentru:										
(a) Studiul după manual, suport de curs, bibliografie şi notiţe										32
(b) Documentare suplimentară în bibliotecă, pe platforme electronice de specialitate şi pe teren										18
(c) Pregătire seminarii / laboratoare, teme, referate, portofolii şi eseuri										43
(d) Tutoriat										0
(e) Examinări										2
(f) Alte activităţi:										0

3.4 Total ore studiu individual (suma (3.3(a)...3.3(f)))	94
3.5 Total ore pe semestru (3.2+3.4)	150
3.6 Numărul de credite	6

4. Precondiţii (acolo unde este cazul)

4.1 de curriculum	Baze de date
4.2 de competenţe	Statistică și probabilități

5. Condiţii (acolo unde este cazul)

5.1. de desfăşurare a cursului	Tabla, proiector, calculatoare
5.2. de desfăşurare a seminarului / laboratorului / proiectului	Tabla, proiector, calculatoare

6. Competenţele specifice acumulate

6.1 Competenţe profesionale	C2. Investigarea și analiza situațiilor de criminalitate informatică și a software-ului malițios prin metode avansate de tip inginerie inversă și monitorizare comportament • C2.1 – Cunoașterea aprofundată a clasificării, caracteristicilor și particularităților aferente diferitelor tipuri de atacuri informatice și softurilor malițioase • C2.3 – Capacitatea de a face corelări și de a putea identifica obiecte potențial malițioase chiar dacă nu se poate analiza complet obiectul respectiv • C2.4 – Determinarea limitărilor teoretice și practice oferite de diverse

	metode de automatizare a analizei software-ului malițios. Propunerea de alternative mai bune unde este posibil • C2.5 – Elaborarea unor noi clase de atacuri sau software malițios, și propunerea unor noi proprietăți potrivite clasificării codurilor malițioase C4. Proiectarea și dezvoltarea de software cu un înalt grad de securitate, de soluții și unelte de securitate • C4.2 – Identificarea de noi scenarii în care este nevoie de introducerea unei soluții de securitate sau utilizarea unei unelte de securitate. Analiza soluțiilor de securitate propuse și compararea lor cu cele cunoscute anterior • C4.3 – Dezvoltarea unor module software complexe respectând principiile metodologiilor de dezvoltare corectă a unui software din perspectiva securității. Dezvoltarea unor utilitare de analiză sau de validare a securității • C4.5 – Dezvoltarea unor module software sau a unor utilitare care să ajute la asigurarea unui înalt grad de securitate. Propunerea unor scenarii și modalități de testare a unor proiecte existente, cu scopul de a verifica și asigura calitatea lor din punctul de vedere al securității C5. Rezolvarea corectă și eficientă a unor probleme complexe de securitate informatică din lumea reală. Operarea cu metode și modele matematice, tehnici și tehnologii aferente inginerești și informatice specifice domeniului securității informațiilor și sistemelor de calcul • C5.1 – Cunoașterea legăturilor dintre securitatea informațiilor și lumea reală. Cunoașterea elementelor matematice care stau la baza elementelor de securitate • C5.2 – Analiza și interpretarea de situații noi complexe din lumea reală, prin prisma cunoștințelor fundamentale din domeniul securității informațiilor și sistemelor de calcul. Identificarea și corelarea unor soluții similare cu cele cunoscute, precum și plasarea corectă a ideilor noi în domeniul cercetării și dezvoltării de soluții de securitate informatice • C5.3 – Aplicarea unor modele matematice și informatice teoretice sau cu o arie mai generală de aplicabilitate pentru a analiza, evalua și rezolva probleme diverse de securitate/confidențialitate din lumea reală • C5.5 – Realizarea de activități de cercetare cu finalitate practică demonstrată prin prototipuri software și/sau hardware funcționale, cu aplicabilitate în domeniul securității informațiilor și sistemelor de calcul
6.2 Competenţe transversale	N/A

7. Obiectivele disciplinei

7.1 Obiectivul general al disciplinei	Dobândirea deprinderii de a lucra cu colecții masive de date. Ținând cont că numărul de programe malițioase este în continuă creștere, se dorește ca studenții să fie capabili să le gestioneze, clasifice și să găsească noi modele pentru detecția acestora.
7.2 Obiectivele specifice	• Dobândirea abilității de a utiliza limbaje de scripting și baze de date pentru a manipula colecții masive de date • Înțelegerea paradigmei Map-Reduce și abilitatea de a proiecta și implementa sisteme distribuite • Cunoașterea unor algoritmi și tehnici pentru căutarea datelor în colecții masive • Deprinderea unor modele pentru clasificarea și învățarea automată a regulilor de detecție a programelor malițioase

8. Conţinuturi

8.1 Curs	Nr.ore	Metode de predare	Observații
Introducere în limbaje de scripting: limbajul Python	2	Expunere la tablă, prezentare cu video-proiectorul, discuții	
Procesarea datelor în Python	2		
Baze de date relaționale și nerelaționale: proprietățile ACID,	2		

algebra relațională, teorema CAP			
Paradigma Map-Reduce	2		
Analiza complexității algoritmilor Map-Reduce	2		
Tehnici simple de căutare în colecții masive: indexare, hashing	2		
Tehnici avansate de căutare în colecții masive:căutarea elementelor similare, identificarea similarităților între programe	2		
Tehnici avansate de căutare în colecții masive: reversed index, locality-sensitive hashing	2		
Analiza link-urilor: algoritmul PageRank, tehnici de SEO	2		
Tehnici de clustering: K-means, clustering ierarhic	2		
Tehnici avansate de clustering penru colecții masive	2		
Construcția de modele pentru predicții: regresie liniară, regresie logistică, arbori de decizie	2		
Clasificatori avansați: perceptronul, Support Vector Machines	2		
Reducerea dimensionalității	2		

Bibliografie *(bibliografia minimală a disciplinei conținând cel puțin o lucrare bibliografică de referință a disciplinei, care există la dispoziţia studenţilor într-un număr de exemplare corespunzător)*

1. Mining of Massive Datasets (Rajarman, Anand – 2011 – Cambridge)
2. Pattern Recognition and Machine Learning (Bishop, Christopher – 2007 – Springer)
3. MongoDB: The Definitive Guide (Chodorow, Kristina – 2013 – O'Reilly) (2nd ed)
4. Data Science for Business: What you need to know about data mining and data-analytic thinking (Provost, Foster – 2013 – O'Reilly)
5. Learning Python (Lutz, Mark – 2013 – O'Reilly) (5th ed)
6. Diferite articole

8.2 Aplicații (seminar/laborator/proiect)*	Nr.ore	Metode de predare	Observații
Introducere în limbajul Python	2	Scurte expuneri la tablă, tutoriale, ghiduri de lucru, demonstrații *live*, explicații suplimentare, discuții, propunerea spre rezolvare a unor probleme de diferite tipuri și grade de complexitate	
Structuri de date în Python	2		
Biblioteci de funcții specifice pentru operarea cu colectii de date	2		
Extragerea de trăsături pentru identificarea programelor potențial malițioase	2		
Stocarea și accesarea colecțiilor de date: baze de date, utilizarea indecșilor	2		
Algoritmi de tip Map-Reduce pentru procesarea colecțiilor de fișiere potențial malițioase	2		
Calculul similarității între programe	2		
Construcția unui reversed index, folosind tehnica Map-Reduce	2		
Regăsirea aplicațiilor similare în colecții masive: locality-sensitive hashing	2		
Tehnici de clustering pentru identificarea familiilor de malware, partea 1	2		
Tehnici de clustering pentru identificarea familiilor de malware, partea 2	2		
Clasificatori pentru malware și spam, partea 1	2		
Clasificatori pentru malware și spam, partea 2	2		
Evaluare și verificare	2		

Bibliografie *(bibliografia minimală pentru aplicații conținând cel puțin o lucrare bibliografică de referință a disciplinei care există la dispoziţia studenţilor într-un număr de exemplare corespunzător)*

1. Mining of Massive Datasets (Rajarman, Anand – 2011 – Cambridge)
2. Pattern Recognition and Machine Learning (Bishop, Christopher – 2007 – Springer)
3. MongoDB: The Definitive Guide (Chodorow, Kristina – 2013 – O'Reilly) (2nd ed)
4. Data Science for Business: What you need to know about data mining and data-analytic thinking (Provost, Foster – 2013 – O'Reilly)
5. Learning Python (Lutz, Mark – 2013 – O'Reilly) (5th ed)
6. Diferite articole

** Se vor preciza, după caz: tematica seminariilor, lucrările de laborator, tematica și etapele proiectului.*

9. Coroborarea conţinuturilor disciplinei cu aşteptările reprezentanţilor comunităţii epistemice, asociaţiilor profesionale şi angajatorilor reprezentativi din domeniul aferent programului

Se realizează prin discuții periodice cu reprezentanți ai angajatorilor semnificativi, în special cei care angajează pe proiecte din domeniul securității informațiilor.
Cursuri cu topic-uri legate de big data sunt prezente în cadrul multor alte masterate, însă puține dintre ele se focalizează și pe domeniul securității calculatoarelor și a informațiilor. Oricum, atât clasificarea software-ului malițios, cât și a spam-urilor reprezintă în practică colecții masive de date, a căror prelucrare necesită metode de învățare automată și prelucrare a datelor massive. Exemple ar fi:

- *Big Data*, Masters in Computer and Information Security, University of Liverpool, UK http://www.liv.ac.uk/study/online/programmes/information-technology/msc-computer-and-information-security/module-details/
- *Big Data Management and Security*, Graduate Certificate Program, Missouri University of Science and Technology, USA, http://dce.mst.edu/credit/certificates/bigdatamanagementandsecurity/
- CS246, *Mining Massive Data Sets*, Stanford, http://web.stanford.edu/class/cs246/
  CSE 599, *Machine Learning for Big Data*, Computer Science & Engineering, University of Washington

Există numeroase alte programe de masterat care se specializează pe big data și business analytics, care predau metode și tehnici care se pot aplica cu succes și în cadrul analizei datelor de securitate.

10. Evaluare

Tip activitate	Criterii de evaluare	Metode de evaluare	Pondere din nota finală
Curs	Abilitatea de rezolvare a unor probleme specifice domeniului Prezență, (inter)activitate în timpul orelor de curs	Examen de tip grilă, (20%), și/sau prezentarea unei teme de cercetare din domeniul cursului (30%)	50%
Seminar			
Laborator	Abilitatea de rezolvare a unor probleme specifice domeniului	Realizarea activităților de laborator și rezolvarea temelor de casă și/sau a unor probleme în cadrul unui examen practic	50%
Proiect			

Standard minim de performanţă

Curs. Prezența la **minim 50%** din orele de curs, pentru admiterea la examenul final.
Aplicații. Prezența la laborator **obligatorie 100%** (2 laboratoare se pot recupera în timpul semestrului, iar altele 2 în sesiunile de restanțe) pentru admiterea la examenul final.
La finalul cursului, studenții trebuie să fie capabili să manipuleze colecții mari de date, atât nestructurate cât și structurate în baze de date, folosind algoritmi secvențial și distribuiți, de tip Map-Reduce. Principalele operații pe care studenții trebuie să demonstreze că le-au deprins sunt căutarea în colecții masive, clasificarea și clusterizarea elementelor, respectiv construirea de modele predictive.

Data completării:	**Titulari**	**Titlu Prenume NUME**	**Semnătura**
	Curs	Conf.dr.ing. Camelia LEMNARU	
	Aplicații	Conf.dr.ing. Ciprian OPRIȘA	

Data avizării în Consiliul Departamentului Calculatoare	Director Departament Prof.dr.ing. Rodica Potolea
Data aprobării în Consiliul Facultății de Automatică și Calculatoare	Decan Prof.dr.ing. Liviu Miclea