MD-2045, CHIŞINĂU, STR. STUDENȚILOR, 7/9, TEL: 022 50-99-01 | FAX: 022 50-99-05, [www.utm.md](http://www.utm.md)

**BIG DATA**

1. **Date despre unitatea de curs/modul**

|  |  |
| --- | --- |
| **Facultatea** | Calculatoare, Informatică și Microelectronică |
| **Catedra/departamentul** | Informatică si Ingineria Sistemelor |
| **Ciclul de studii** | Studii superioare de licenţă, CICLUL I |
| **Programul de studiu** | 0612.2 Managementul Informaţiei |
| **Anul de studiu** | **Semestrul** | **Tip de evaluare** | **Categoria formativă** | **Categoria de opţionalitate** | **Credite ECTS** |
| III (învăţământ cu frecvenţă); | V | E, PA | S – unitate de curs de specialitate | A - unitate de curs la alegere | 4 |

1. **Timpul total estimat**

|  |  |
| --- | --- |
| Total ore în planul de învăţământ | Din care |
| Ore auditoriale | Lucrul individual |
| Curs | Lucrări practice | Proiect de an | Studiul materialului teoretic | Pregătire aplicaţii |
| 120 | 30 | 30 | +  | 30 | 30 |

1. **Precondiţii de acces la unitatea de curs/modul**

|  |  |
| --- | --- |
| Conform planului de învăţământ | Studiul se bazează pe următoarele discipline: Matematică discretă, Algebră, Teorie probabilități și statistici matematice, Metode de programare, Metode de analiză a datelor. Aditional, studenții trebuie să cunoască, structura calculatorului, SO Windows 10, limbajele de programare C#, C++, HTML, PHP, Python, utilizarea unui redactor de texte. Cunoştinţe avansate despre sisteme moderne de Gestiune a Bazelor de Date, cum ar fi: MySQL, MS SQL Server 2019, CMS-uri Wordpress, framework-uri Laravel, Code Igniter, Symphony ș.a., abilităţi de a crea aplicaţii DeskTop sau Web, Front End și Back End, utilizând Baze de Date Operaționale, precum și modelarea multidimensională a datelor, Depozitele de date, tehnologia OLAP, Instrumente BI, (VS, Power BI), tehnologia Data Mining, lucrul cu tehnologia cloud, platformele Azure, AWS (Amazone Web Service) și Google Cloud. |
| Conform competenţelor | Studenții trebuie să aibă următoarele cunoștințe și competențe pentru a stăpâni disciplina academică “Big Data”: capacitatea de a utiliza corect TIC în rezolvarea problemelor profesionale, cu caracter aplicativ, folosind abilitățile de modelare matematică ale fenomenelor social-economice, ale mediului ambiat, din domeniul tehnic, precum și din domeniile: financiar-bancar, învățământ, medicină, asigurări, marketing etc, abilități formate la disciplinele menționate mai sus. Rezultatele principale ale disciplinei “Big Data” pot fi utilizate în viitor la studierea disciplinelor: “Analiza datelor”, “Metode de protecție a informațiilor”, “Știinșa Datelor”, “Analitica Big Data”, “Informatica in afaceri”, “Procesarea Big Data”, “Analiza Big data”, “Tehnologii Big Data”,  |

1. **Condiţii de desfăşurare a procesului educaţional pentru**

|  |  |
| --- | --- |
| Curs | Pentru prezentarea materialului teoretic sala de curs trebuie sa dispuna de un proiector şi calculator. Nu vor fi tolerate întârzierile studenţilor, precum şi convorbirile telefonice în timpul cursului. |
| Laborator/seminar/practice | Studenţii vor perfecta rapoarte conform condiţiilor prevazute de indicaţiile metodice. Termenul de predare a lucrării de laborator/practice, individual sau de grup, se prezintă la următoarea ședință de lucru, în format electronic. Pentru predarea cu întârziere a lucrului efectuat, aceasta se depunctează cu 1pct./săptămână de întârziere. |

1. **Competenţe specifice acumulate**

|  |  |
| --- | --- |
| Competenţe profesionale | **CP1. Managementul nivelului de servicii (A.2)\*\**** K2 Cum se compară și se interpreteze datele de management.

**CP2. Proiectarea și dezvoltarea aplicațiilor (B.1)*** K4 Tehnologiile de ultimă oră.
* K5 Limbaje de programare.
* K6 Baze de date (DBMS).
* K7 Sisteme de operare și platforme software.
* K8 Mediul de dezvoltare integrat (IDE - integrated development environment).

**CP3. Integrarea componentelor (B.2)*** K2 Impactul integrării unui sistem asupra organizației sau a sistemului existent.
* K4 Tehnici de testare a integrării.
* K5 Instrumentele de dezvoltare (ex. mediul de dezvoltare, gestionare, control al modificărilor și accesul la codul sursă).
* K6 Bune practici de design

**CP5. Furnizarea de servicii (C.3)*** K1 Modul de interpretare a cerințelor privind prestarea de servicii IT.
* K2 Cele mai bune practici și standarde pentru prestarea serviciilor informatice.
* K6 Specificul tehnologiilor legate de web, cloud și instrumente mobile.

**CP6. Managementul informațiilor și a cunoștințelor (D.10)*** K1 Metodele de analiză a informațiilor și a proceselor de business.
* K3 Provocările legate de dimensiunea masivelor de date (Big Data).
* K4 Provocările legate de date nestructurate (de exemplu, Data Analytics).

**CP7. Identificarea nevoilor (D.11)** * K2 Nevoie de afaceri.
* K3 Procesele și structura organizației.
* K4 Tehnici de analiză a nevoilor clienților.

**CP8. Marketing digital (D.12)*** K1 Strategii de marketing.
* K2 Tehnologiile web.
* K9 Probleme/cerințe legale

**CP9. Managementul relațiilor cu clienții (E.4)** * K1 Procesele de organizare, inclusiv luarea deciziilor, bugetele și structura de management.
 |
| Competenţe transversale | **CTL1 Autonomie şi responsabilitate** Demonstrează executarea responsabilă a sarcinilor profesionale în condiții de autonomie.**CTL2 Interacțiune socială**Execută rolurile şi activitățile specifice muncii în echipă şi distribuie sarcinile între membri pe niveluri subordonate.**CTL3 Dezvoltare personală şi profesional**Conștientizează nevoia de formare continuă cu utilizarea eficientă a resurselor şi tehnicilor de învățare pentru dezvoltarea personală şi profesională |

1. **Obiectivele unităţii de curs/modulului**

|  |  |
| --- | --- |
| Obiectivul general | familiarizarea, iniţierea şi instruirea studenţilor pentru a înțelege și utiliza în practică care este esența și specificul cursului “Big Data”, cum se poziționează în acest proces utilizarea practică a cunoștințelor și abilităților obținute la cursurile precedente ce l-au precedat, „***Baze de Date***” şi “***Managementul Bazelor de date***”, cum se aplică pentru ***managementul Big Data***, Bazele de Date Operaționale - BDO, Depozitele de Date - DWH și sistemele moderne de gestiune a lor, precum şi **să dezvolte deprinderi, abilități practice și să** **asigure dobândirea de competenţe, atât pe dimensiunea analitică (modelare statistică, metode predictive, tehnici de învăţare automată) cât şi pe cea tehnologică (limbaje de programare, arhitecturi şi platforme).** ce le-ar permite studenților, viitorii specialiști, să ***aplice cunoştinţele acumulate, de-a lungul cursurilor menționate***, la soluţionarea celor mai diverse probleme din economia naţională, utilizând simularea proceselor şi fenomenelor din domenii de activitate proxime viitorilor specialişti în aplicarea TIC, instruiţi în cadrul ***facultăţii Calculatoare Informatică şi Microelectronică***, la programul de studii - ***0612.2 Managementul Informaţiei*** |
| Obiectivele specifice | *La nivel de cunoștinţe:** Capacitatea de a opera cu concepte fundamentale din domeniul modelării matematice şi analizei statistice precum şi abilitatea de a le folosi în contexte practice
* Capacitatea de a identifica, implementa şi utiliza algoritmi de extragere a modelelor din date folosind metode statistice şi tehnici de învăţare automată.
* Capacitatea de a înţelege şi a aplica principiile procesării distribuite a datelor şi a utilizării arhitecturilor de calcul de înaltă performanţă.
* Abilitatea de a utiliza platforme şi tehnologii specifice prelucrării unor volume mari de date şi de a implementa aplicaţii scalabile.
* Capacitatea de a adapta soluţii bazate pe abordări de tip data-driven pentru probleme specifice unui anumit domeniu de aplicabilitate.
* Capacitatea de a planifica şi organiza activitatea de cercetare într-un domeniu prin organizarea lucrului în grup; înţelegerea responsabilităţilor şi respectarea regulilor de etică profesională şi a celor de securitate şi confidenţialitate a datelor.
* Utilizarea unor metode şi tehnici eficiente de învăţare, informare şi specializare; abilitatatea de a selecta şi analiza critic resursele bibliografice; asumarea unui plan de dezvoltare profesională continuă prin actualizarea permanentă a cunoştinţelor şi abilităţilor în corelare cu progresele ştiinţifice şi tehnologice înregistrate în domeniu

*La nivel de abilităţi:** să elaboreze strategii de date Big Data, reieșind din contextul entității ce le utilizează, scopul și obiectivele formulate pentru dezvoltarea ei și generarea de valoare adugată;
* să dezvolte fluxuri și conducte de date necesare procesului de management al “Big Data”, pentru facilitarea activităților de management informational necesar procesului decisional al entității;
* să analizeze utilizarea “Big Data” din punctul de vedere al eficienţei, utilizarii datelor, atunci când este vorba de un Proiect informațional concret, ce tine de procesul decizional al unei entitati, suport al caruia “Big Data” se prezuma a fi;
* să utilizeze lp. Python, platformele Power BI, VS, tehnologii cloud si platforme cloud cum ar fi: AWS, Azure, Google Cloud, pentru Analiza si Analitica Big Data.
 |

1. **Conţinutul unităţii de curs/modulului**

|  |  |
| --- | --- |
| **Tematica activităţilor didactice** | Numărul de ore |
| învăţământ cu frecvenţă | învăţământ cu frecvenţă redusă |
| **Tematica prelegerilor** |
| **TP1.**  **Introducere în Big Data și Infrastructura sa.** | **2** |  |
| **TP2. Infrastructura și Arhitectura Big Data**. Componente hardware/software, sisteme de stocare, soluții scalabile. | **2** |  |
| **TP3. Tehnologii Specifice Big Data.** Tehnologii fundamentale: Hadoop, Spark, Hive, Pig, NoSQL și Baze de date distribuite. | **2** |  |
| **TP4. Programare Distribuită cu MapReduce și Hadoop.**Paradigma de programare distribuită, pentru a dezvoltarea conceptului de Hadoop și arhitectura sa. | **4** |  |
| **TP5. Procesare Avansată cu Apache Spark.** | **2** |  |
| Spark Core, RDD, DataFrames, optimizări. | **2** |  |
| **TP6**. **Fluxuri de Date și Procesare în Timp Real cu Kafka și Flink. Spark Streaming**. | **2** |  |
| **TP7. Baze de Date NoSQL: HBase și Cassandra.**  | **2** |  |
| **TP8. Inteligența Artificială în Big Data** Algoritmi AI, Machine Learning, Deep Learning în contextul Big Data. | **2** |  |
| **TP9. Vizualizarea Avansată a Datelor cu Baze de Date Graf Neo4j** | **4** |  |
| **TP10. Big Data, IoT** **și** **IIoT**. Captarea și agregarea datelor din IoT și IioT; Analiza și prelucrarea lor.Interoperabilitate și Integrarea surselor multiple de date.Provocările legate de stocare și procesare în timp real.Datele IoT, IIoT: analiza și prelucrarea lor; Integrarea cu platforme Big Data | **4** |  |
| **TP11. Societatea și Economia bazată pe date.** | **2** |  |
| **Total prelegeri:** | **30** |  |

|  |  |
| --- | --- |
| Tematica activităţilor didactice | Numărul de ore |
| învăţământ cu frecvenţă | învăţământ cu frecvenţă redusă |
| **Tematica lucrărilor de laborator** |
| **Lucrare de laborator nr. 1.*** Explorarea seturilor de date mari și vizualizarea distribuțiilor inițiale.
* Configurarea unei infrastructuri Big Data în mediu virtual.
* *Instalarea și configurarea unui cluster Hadoop.*
* *Importul unui set de date.*
* *Scrierea și rularea programelor MapReduce pentru diverse scenarii de prelucrare a datelor.*
* *Optimizarea performanței joburilor MapReduce.*
 | **10** |  |
| **Lucrare de laborator nr. 2.*** *Crearea de aplicații Spark pentru prelucrarea și analiza datelor.*
* *Utilizarea SparkSQL pentru interogarea datelor.*
* Implementarea unei soluții de procesare a fluxurilor de date în timp real.
* Implementarea unei soluții de procesare a fluxurilor de date în timp real.
* *Dezvoltarea și implementarea de pipeline-uri ETL/ELT*.
* *Configurarea și utilizarea instrumentelor de guvernare a datelor.*
 | **10** |  |
| **Lucrare de laborator nr. 3.*** Crearea și interogarea bazelor de date graf în Neo4j.
* Antrenarea și evaluarea modelelor de machine learning pe date reale.
* Utilizarea pipeline-urilor pentru fluxuri de lucru de ML.
* Dezvoltarea de vizualizări complexe și interactive.
* Construirea unui model de învățare automată pe un set de date mari.
* Analiza datelor economice din surse multiple.
* Analiza economiei digitale și rolul Big Data.
* Analiza și construirea unui model de învățare automată pe un set de date.
 | **10** |  |
| **Total lucrări de laborator:** | **30** |  |

1. **Referinţe bibliografice**

|  |  |
| --- | --- |
| Princi-pale | 1. **Big Data Analytics with Spark,** Copyright © 2015 by Mohammed Guller, ISBN-13 (pbk): 978-1-4842-0965-3, ISBN-13 (electronic): 978-1-4842-0964-6, [www.apress.com/source-code/](http://www.apress.com/source-code/), OR [www.it-ebooks.info](http://www.it-ebooks.info)
2. **Thomas Erl, Wajid Khattak and Paul Buhler,** **Big Data Fundamentals, Concepts, Drivers & Techniques,** Copyright © 2016 Arcitura Education Inc., ISBN-13: 978-0-13-429107-9
3. **Raul Estrada Isaac Ruiz,** ***Big Data SMACK*: *A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka,*** Mexico City Mexico City, ISBN-13 (pbk): 978-1-4842-2174-7 ISBN-13 (electronic): 978-1-4842-2175-4, DOI 10.1007/978-1-4842-2175-4, Library of Congress Control Number: 2016954634, Copyright © 2016 by Raul Estrada and Isaac Ruiz
4. **Wes McKinney, Python for Data Analysis,**  Copyright © 2013 Wes McKinney. All rights reserved. Printed in the United States of America. Published by O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.
5. **John Wiley & Sons,** **Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data,** 10475 Crosspoint Boulevard, Indianapolis, IN 46256, [www.wiley.com](http://www.wiley.com) Copyright © 2015 by John Wiley & Sons, Inc., Indianapolis, Indiana, Published simultaneously in Canada ISBN: 978-1-118-87613-8, ISBN: 978-1-118-87622-0 (ebk) ISBN: 978-1-118-87605-3 (ebk)
6. **Peters Morgan,** **Data analysis from scratch with Python. Step by step guide** © Copyright 2016 by AI Sciences LLC, All rights reserved. First Printing, 2016, Edited by Davies Company Ebook Converted and Cover by Pixels Studio Publised by AI Sciences LLC, ISBN-13: 978-1721942817, ISBN-10: 1721942815
7. **Building Big Data and Analytics Solutions in the Cloud** International Technical Support Organization December 2014
8. **Big Data Principles and Paradigms, Edited by Rajkumar Buyya The University of Melbourne and Manjrasoft Pty Ltd, Australia Rodrigo N. Calheiros** The University of Melbourne, Australia Amir Vahid Dastjerdi The University of Melbourne, Australia, Morgan Kaufmann is an imprint of Elsevier 50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA Copyright © 2016 Elsevier Inc. All rights reserved.496 pag
9. **10 Best Big Data Books in 2023 [Learn Big Data ASAP]**

 <https://azure.microsoft.com/en-us/products>  <https://realtoughcandy.com/best-big-data-books/> 1. **20 Best Big Data Books for Beginners**

 <https://bookauthority.org/books/beginner-big-data-books> 1. **Must Read Books for Beginners on Big Data, Hadoop and Apache Spark**

 <https://www.analyticsvidhya.com/blog/2015/10/books-big-data-hadoop-apache-spark/>  <https://dokumen.pub/big-data-concepts-technology-and-architecture-9781119701828.html>1. **MIcrosoft Aure for students**

 <https://engineering.buffalo.edu/computer-science-engineering/information-for-students/information-technology/software-distribution/microsoft-azure-for-students.html>  <https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/azure-students-program>  <https://azure.microsoft.com/en-us/resources/students?activetab=pivot:githubtab>  <https://learn.microsoft.com/en-us/azure/education-hub/azure-dev-tools-teaching/program-faq> <https://www.techopedia.com/definition/26434/azure-service-platform> 1. **Big Data Analytics** - **Statistical Methods.** [**https://www.tutorialspoint.com/big\_data\_analytics/statistical\_methods.htm**](https://www.tutorialspoint.com/big_data_analytics/statistical_methods.htm)
2. **Майер-Шенбергер, В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим** / Виктор Майер-Шенбергер, Кеннет Кукьер ; пер. с англ. Инны Гайдюк. — М. : Манн, Иванов и Фербер, 2014. — 240
3. **Гусева Е. Н. Теория вероятностей и математическая статистика** : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Москва : ФЛИНТА , 2016 - 220 – Режим доступа: <http://biblioclub.ru/index.php?page=book_red&id=83543>
4. **Гутова С. Г., Алтемерова О. А. Теория вероятностей и математическая статистика:** Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Кемерово : Кемеровский государственный университет , 2016 - 216 - Режим доступа: <http://biblioclub.ru/index.php>? page=book\_red&id=481538
5. **Кочетков Евгений Семенович. Теория вероятностей и математическая статистика:** Учебник [Электронный ресурс] : Форум , 2018 - 240 - Режим доступа: http://znanium.com/go.php?id=944923
6. **Миркин Б. Г. Введение в анализ данных.** Учебник и практикум[Электронный ресурс] : М.:Издательство Юрайт , 2019 - 174 - Режим доступа: https://biblioonline.ru/book/vvedenie-v-analiz-dannyh-432851
 |
| Supli-mentare | 1. **Белько Иван Васильевич.** Теория вероятностей, математическая статистика,

математическое программирование : Учебное пособие [Электронный ресурс] , 2016 - 299 -Режим доступа: <http://znanium.com/go.php?id=542521>1. **Волкова Полина Андреевна.** Статистическая обработка данных в учебно-

исследовательских работах : Учебное пособие [Электронный ресурс] : Форум , 2019 - 96 - Режим доступа: <http://znanium.com/go.php?id=1030246>1. **Непомнящая Наталья Васильевна.** Статистика: общая теория статистики,

экономическая статистика. Практикум : Учебное пособие [Электронный ресурс] , 2015 - 376 -Режим доступа: <http://znanium.com/go.php?id=549841>1. **Основы теории вероятностей и математической статистики** [Электронный ресурс] , 2018 - 90 - Режим доступа: https://lib.rucont.ru/efd/684477
 |

1. **Evaluare**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Periodică** | **Curentă** | **Studiu individual** | **Proiect/teză** | **Examen**  |
| **EP 1** | **EP 2** |
| **15%** | **15%** | **15%** | **---** | **15%** | **40%** |
| **Standard minim de performanţă**Prezenţa şi activitatea la prelegeri şi lucrările practice;Obţinerea notei minime de „5” la fiecare dintre atestări şi lucrări practice individuale sau de grup;Demonstrarea în lucrarea de examinare finală a cunoştinţelor teoretice şi competențelor practice necesare pentru utilizarea Big Data în cadrul dezvoltării unui Proiect concret. |