Yuriy Fedkovych Chernivtsi National University

(full name of the higher education institution)

Educational and Scientific Institute of Physical, Technical and Computer Sciences

(name of the institute/faculty)

Department of Computer Systems and Networks

(name of the department)

SILABUS

of academic discipline

IoT & IoE Technologies for Big Data Analysis

(indicate the name of the discipline (in foreign language, if the discipline is taught in a foreign language))

mandatory

(mandatory or optional)

Educational and scientific programme - *Computer Engineering of*

Internet of Things and Cyber-Physical Systems Technologies

Speciality *123 - Computer engineering*

(code and name of speciality)

Field of knowledge *12 - Information technology*

(indicate: first (bachelor's)/second (master's)/third (educational and scientific))

Educational and Scientific Institute of Physical, Technical and Computer Sciences

(name of the faculty/institute where specialists are trained under the specified educational and professional programme)

Languages of instruction: *Ukrainian, English*

(language in which the discipline is taught)

Developer: Yuliya Tanasyuk, Associate Professor of the Department of Cjvputer Systems and Networks, PhD in Physics and Mathematics

(indicate authors (lecturer (ив)), their positions, academic degrees, academic titles)

Profile of the teacher(s) https://csn.chnu.edu.ua/employees/tanasyuk-yuliya-volodymyrivna/

Contact phone. +(38) 0372 50 94 32.

E-mail: y.tanasyuk@chnu.edu.ua

Course page in Moodle https://moodle.chnu.edu.ua/course/view.php?id=5098

Consultations *on-line: Monday from 14.40 to 16.00*

1. Summary of the discipline

The study of the discipline "IoT & IoE Technologies for Big Data Analysis" allows students to develop the competencies necessary to solve practical problems of professional and scientific activities related to the processing and analysis of big data generated by devices in the Internet of Things.

2. The purpose of the discipline "IoT & IoE Technologies for Big Data Analysis": to develop students' ability to apply software methods and data processing tools for analysing big data and making management decisions in various fields of science and business.

2.1. Objectives of the discipline:

- Explain how organisations can extract information and generate new insights from IoT data.
- Understand and implement the stages of the big data analytics lifecycle.
- Explain the different types of data analytics: descriptive, predictive, and predictive.
- Use Python to create a process for retrieving, manipulating, and visualising data from IoT sources.
- Apply intelligence analysis to identify emerging trends.
- Explain the purpose of machine learning.
- Be able to represent and transfer different types of data.
- Describe the evolution of manipulation technologies from SQL to NoSQL.
- Understand and explain the development of modern data centres and distributed scalable Big Data platforms such as Apache Hadoop.

3. Prerequisites. To successfully master the material, students must first study the following disciplines: Fundamentals of Algorithmisation, Database Organisation, Probability Theory and Mathematical Statistics, Algorithms and Methods of Computing, Computer Systems, OOP, Object Communication Devices, Microcontrollers, Fundamentals of IoT & IoE, Computer Systems of AI, Cloud Computing.

4. Learning **outcomes**

As a result of studying the discipline, students should:

Know:

- architectural models of Big Data;
- virtualisation technologies and container technologies for executing software codeon the server;
- Hadoop Big Data, MapReduce, HDFS, Kafka, Cassandra, Spark technologies;
- the opportunities provided by the Python programming language for big dataanalysis;
- technologies for importing data from the Internet and tools for correlation analysisin Pandas;
- data management and processing tools available in SQL and NoSQL databases;
- technologies for regression analysis of big data using Python;

- data visualisation technologies Tableau, Pyplot, Plotly, Scikit-Learn, Seaborn, Folium, Leaflet.js Pandas, and Matplotlib;
- methods and tools for analysing big data using cloud, fog and edge computing;
- measures for organising big data processing and visualisation based on RaspberryPi.

Be able to:

- use IoT technologies and software tools to process and analyse big data;
- using the Python programming language to create a full cycle of data extraction, manipulation and presentation from various sources;
- apply exploratory data analysis to generate decisions based on the data;
- present and visualise large amounts of data;
- process poorly structured and diverse data from various sources using Python andNoSQL databases.

Acquire competences:

General competences (GC):

GC1. Ability to adapt and act in a new situation.

GC2. Ability to think abstractly, analyse and synthesise.

GC4. Ability to search, process and analyse information from various sources.

GC6. Ability to identify, formulate and solve problems.

GC7. Ability to make informed decisions.

GC8. Ability to communicate in a foreign language.

Special (professional) competences (SC):

SC2. Ability to develop algorithmic and software, components of computer systemsand networks, Internet applications, cyber-physical systems using modern programming methods and languages, as well as design automation tools and systems.

SC4. Ability to build and research models of computer systems and networks.

SC6. Ability to use and implement new technologies, including smart, mobile,green and secure computing technologies, to participate in the modernisation andreconstruction of computer systems and networks, various embedded and distributed applications, in particular to improve their efficiency.

SC9. Ability to present the results of own research and/or development in the formof presentations, scientific and technical reports, articles and papers at scientific and technical conferences.

SC10. Ability to identify, classify and describe the operation of software and hardware, computer systems, networks and their components;

SC11. Ability to choose effective methods for solving complex computer engineering problems, critically evaluate the results obtained and justify decisions.

SC12. Ability to use methods of analysis, identification and synthesis of computer systems and networks, cyber-physical systems, Internet of Things and IT infrastructures.

Programme learning outcomes (PLOs):

PLO1. Apply general approaches to cognition, methods of mathematics, natural and engineering sciences to solve complex computer engineering problems.

PLO2. Find the necessary data, analyse and evaluate it.

PLO6. Analyse issues, identify and formulate specific problems that need to be solved, choose effective methods of solving them.

PLO8. Apply knowledge of technical characteristics, design features, purpose and rules of operation of software and hardware of computer systems and networks to solve complex problems of computer engineering and related problems.

PLO9. Develop software for embedded and distributed applications, mobile and hybrid systems.

PLO10. Search for information in various sources to solve computer engineering problems, analyse and evaluate this information.

PLO12. Communicate fluently orally and in writing in Ukrainian and one of the foreign languages (English, German, Italian, French, Spanish) when discussing professional issues, research and innovation in the field of information technology.

PLO14. Plan and carry out scientific research in the field of computer engineering, formulate and test hypotheses, choose methods and tools, analyse results, and justifyconclusions.

PLO15. Conduct research and development in the field of specialised, problem-oriented high-performance computer systems using modern IoT and CFC technologies, mobile and hybrid computing, Big Data analysis, the latest hardware and software solutions based on modern element base, in particular, programmable logic integrated circuits, microcontrollers, microcomputers, multi-core processors.

5. Description of the discipline

5.1. General information

Title of the discipline *IoT & IoE technologies for Big Data Analysis*												
Form of study	**Year of training**	**Semester**	**Quantity**			**Number of hours**						**Type of final control**
			credits	**hours**	**content modules**	**lectures**	**practical**	**seminar**	**laboratory**	**independent work**	**individual tasks**	
Full-time	**2(6)**	**3 (11)**	**4**	**120**	**2**	**15**	**-**	**-**	**30**	**75**	**-**	**Exam**

5.2. Didactic map of the discipline

Theoretical content of the programme of the discipline

Content module 1: ***Sources and types of big data***

Topic 1: Sources of big data in the concept of the Internet of Things. Definition of Big Data.

Topic 2. Fundamentals of data analysis and machine learning.

Content module 2: ***Architectural models of Big Data***

Topic 3. Data classification.

Topic 4. Virtualisation technologies

Topic 5. Distributed platforms for big data processing

Titles of content modules and topics	Number of hours					
	in total	including				
		l	p	lab	it	ind
1	2	3	4	5	6	7
Content module 1: *Sources and types of big data*						
Topic 1.	20	4	-	2	-	14
Topic 2.	28	2	-	10	-	16
Total for M1	48	6	-	12	-	30
Content module 2: *Architectural models of Big Data*						
Topic 3.	19	3	-	4	-	12
Topic 4.	27	3	-	6	-	18
Topic 5.	26	3	-	8	-	15
Total for M 2	72	9		18	-	45
Total hours	120	15	-	30	-	75

5.3. Topics of laboratory classes

№	Name of the topic (task)	Number of hours
1.	Data analysis and visualisation in Python.	4
2.	Correlation analysis in Python.	4
3.	Descriptive statistics in Python.	4
4.	Building linear regression and forecasting.	4
5.	Visualisation of statistical data.	6
6.	NoSQL databases for distributed big dataprocessing	8
	Total:	30

Note. Methodological recommendations and assignments for laboratory work are available on the following Internet resources:
https://moodle.chnu.edu.ua/course/view.php?id=5098
https://skillsforall.com/course/data-analytics-essentials?courseLang=en-US&instance_id=bd4725ae-c054-4bba-8553-e7be9408f90e

Software for performing laboratory work: Google Colab cloud platform, Python programming language, Jupyter Notebook web shell, Tableau online visualisation tool: https://www.tableau.com/products/public/download . IBM Cloudant cloud environment for working with NoSQL databases: cloud.ibm.com.

5.4. Tasks for independent work

№	Topic title	Number of hours
1	Research of open data sources. Download the dataset and save the data in csv format.	6
2	Python and SQLite programming. The purpose of the csvsql utility.	4
3	Statistical approaches to big data analytics. Using Pandas.Importing data from the Internet. Descriptive statistics in Pandas. Tools for correlation analysis in Pandas.	12
4	Regression analysis of data in Python	5
5	Pyplot module. The Plotly tool. Types of data visualisation.	5
6	Hadoop Big Data technologies. Distributed processing with MapReduce. HDFS.	4
7	Kafka: a distributed streaming big data processing platform. Advantages of Cassandra.	8
8	The problem of a computational function. Spark technology. Comparison of Spark and MapReduce.	12
9	Lambda and Kappa big data processing architectures.	8
10	Machine learning.	4
11	Big data protection.	4
12	Ethics of data use	3
	Total:	75

6. Forms and methods of training

Forms of education include problem-based and review lectures, laboratory classes, classes with the use of computer and telecommunications equipment, interactive classes with one student teaching another, video lectures, video conferencing via Google Meet, Cisco Webex, classes using the Moodle e-learning system and the educational portal of the Cisco Network Academies Programme - netacad.com and skillsforall.com.

Methods: problematic presentation of material, partially search and research laboratory workshops, presentations, consultations and discussions, work in the Internet classroom: electronic lectures, laboratory work, distance consultations, etc. aimed at activating and stimulating students' educational and cognitive activities.

Approaches to teaching: student-centred, problem-based, activity-based, communicative, professionally oriented, interdisciplinary approaches are used.

The educational process is carried out during lectures, laboratory classes, independent extracurricular work with the use of modern information teaching tools, software simulators, virtual teamwork environments, and consultations with teachers.

The following **teaching methods are** used to develop **skills and abilities**:

- verbal/verbal *(lecture, explanation, story, conversation, instruction);*
- visual *(observation, illustration, demonstration);*
- practical *(conducting an experiment, practice);*
- explanatory and illustrative, which involves the presentation of ready-made information by the teacher and its assimilation by students*;*
- reproductive *(performing laboratory tasks based on a sample);*
- problem-based presentation of material in laboratory classes.

7. Monitoring and evaluation system

Means of assessing and demonstrating learning outcomes:

- cloud-based design and visualisation environments;
- tasks on laboratory equipment;
- laboratory work;
- tests;
- presentations and justification of the results of completed tasks.

The forms of current control of the level of knowledge are oral and written answers of the student during the defence of completed laboratory works, presentation of the results of practical tasks in the form of a report, the number of points obtained in the performance of test tasks. The final control is carried out by passing the final exam.

7.1. Criteria for assessing learning outcomes in the discipline

The criterion for the successful completion of the final assessment by the student is the achievement of the minimum threshold levels of grades for each planned learning outcome of the discipline.

Assessment scale and criteria: national and ECTS (European Credit Transfer and Accumulation System)

ECTS grades	Criteria	Explanation	Rating on a 100-point scale	Score on the national scale (exam)
A	Excellent level of competence within the mandatory material, with possible minor deficiencies	excellent	**90 - 100**	**excellent**
B	A sufficiently high level of competence within the mandatory material without significant (gross) errors	very good	**80-89**	**good**
C	Overall good level of competence with few errors	good	**70-79**	
D	Mediocre level of competence with a significant number of deficiencies, sufficient for further study or professional activity	satisfactory	**60-69**	**satisfactory**
E	Minimum acceptable level of competence	enough	**50-59**	
FX	Unsatisfactory level of competence, with the possibility of retaking the exam with proper independent revision	(unsatisfactory) with the possibility of retaking	**35-49**	**unsatisfactory**
F	Very poor level of competence, requiring repeated study of the discipline	(unsatisfactory) with mandatory re-study of the discipline	**1-34**	

7.2. Distribution of the maximum possible number of points that students receive for completing all types of learning activities

Current assessment (individual work and independent work)					Data Analytics Essentials	Final control (exam)	Total numberof points
Content module 1		Content module 2					
T1	T2	T3	T4	T5	MC	40	100
6,5	13,5	6,5	6,5	7	20		

The forms of final control of the level of knowledge are tests, laboratory work and quizws in the distance learning system.

Content module 1: ***Sources and types of big data***

T1. Sources of big data in the concept of the Internet of Things. Definition of Big Data. (completion and defence of lab 1 - 10 points). BDA1 - test on the materials of Topic 1 (10 points).

T2. Fundamentals of data analysis and machine learning (completion and defence of lab 2 - 10 points, completion and defence of the Lab 3 - 10 points).

BDA2 - test on the materials of Topic 2 (10 points).

BDA3 - test on the materials of Topic 2 (10 points).

Content module 2: ***Architectural models of Big Data***

T3. Classification of data (completion and defence of lad 4 - 10 points).

BDA4 - test on the materials of Topic 3 (10 points).

T4. Virtualisation technologies (completion and defence of the lab 5 - 10 points).

BDA5 - test based on the materials of Topic 4 (10 points).

T5. Distributed platforms for big data processing (completion and defence of the lab 6 - 10 points).

BDA6 - test based on the materials of Topic 5 (10 points).

MK - the cumulative result of studying the Data Analytics Essentials online course from Cisco Networking Academy (20 points).

Final test (exam) - 40 points.

The total amount of points for the semester is determined by the formula:

$$STP=((LWT+TCT+DAE*0,6)/3) + FE,$$

where **STP** – total points for semester (maximum – 100);

LWT – total points for laboratory works (maximum – 60);

TCT – points for topic control tests (maximum – 60);

DAE - the final weighted score for the online course "Data Analysis Essentials" (maximum 100 points);

FE – final exam total points (maximum – 40).

7.3. Conditions for crediting non-formal education results

According to the Regulations of the CHNU "On Non-Formal Education", a student can receive additional points or be exempted from certain types of work on certain topics if he or she has certificates of non-formal education on issues that are covered by the discipline "IoT & IoE Technologies for Big Data Analysis".

8. Recommended references

8.1. Main (professional)

1. Інтернет речей для індустріальних і гуманітарних застосунків. У трьох томах. Том 1. Основи і технології / За ред. В. С. Харченка. – Міністерство освіти і науки України, Національний аерокосмічний університет ХАІ, 2019. – 605 с.
2. IoT Fundamentals: Big Data & Analytics. URL: https://www.netacad.com/courses/iot/big-data-analytics
3. Data Analytics Essentials. URL: https://skillsforall.com/course/data-analytics-essentials?courseLang=en-US
4. О'Нил К. Big Data. Зброя математичного знищення. – К: Форс, 2019. – 336 с.
5. Провост Ф., Фоусет Т. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані. – К.: Наш Формат, 2019. 400 с.
6. Nokeri T. C. Data Science Solutions with Python. – Apress, 2022. 119 p.

8.2. Supplemental

1. Klosterman S. Data science projects with Python: a case study approach to gaining valuable insights from real data with machine learning? 2nd edition. Packt Publishing Limited. 2021. 432 p.
2. Kats P. Learn Python by building data science applications. 1st edition. Packt Publishing. 2019. 482 p.
3. Ерл Т. Основи Big Data: концепції, алгоритми і технології. – К: Баланс бізнес бус, 2018. – 320 с.
4. Bruce P. Practical statistics for data scientists: 50 essential concepts. 1st edition. O'Relly Media. 2017. 318 p.
5. Muller A.C. Introduction to machine learning with Python: a guide for data scientists. 1st edition. O'Relly Media. 2016. 402 p.

9. Internet resources

1. Big Data: 20 free Big Data sources everyone should know. URL: https://www.smartdatacollective.com/big-data-20-free-big-data-sources-everyone-should-know/
2. Big Data & Analytics Tutorial. URL: https://www.tutorialspoint.com/big_data_tutorials.htm
3. Big Data Tutorial Library. URL: https://data-flair.training/blogs/big-data-tutorials-home/
4. Machine learning. URL: https://www.w3schools.com/python/python_ml_getting_started.asp
5. Python Machine Learning Tutorials. URL: https://realpython.com/tutorials/machine-learning/
6. Data Science Tutorial for Beginners. URL: https://www.simplilearn.com/tutorials/data-science-tutorial