Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Computer Systems Software Department

DATA ANALYSIS

Working program of the academic discipline (Syllabus)

Details of the academic discipline	
Level of Higher Education	***First (bachelor's)***
Field of Study	*12 Information Technologies*
Specialty	*121 Software Engineering*
Education Program	*Software Engineering of Multimedia and Information Retrieval Systems*
Type of Course	*Selective*
Mode of Studies	*full-time*
Year of studies, semester	*3rd year, autumn semester*
ECTS workload	*Lectures: 36 hours, laboratory classes: 18 hours, independent work: 66 hours.*
Testing and assessment	*Assessment, modular control work, calendar control*
Course Schedule	*According to the schedule for the spring semester of the current academic year (rozklad.kpi.ua)*
Language of Instruction	*English*
Course Instructors	Lecturer: *Ph.D., Associate Professor, Onai Mykola* Practical training: *Ph.D., Associate Professor, Onai Mykola*

Program of educational discipline

1. Course description, goals, objectives, and learning outcomes

The study of the discipline "Data Analysis" allows students of higher education to develop the competencies necessary for solving complex problems of professional activity related to the development of software systems for solving typical problems that arise during the analysis of large volumes of accumulated experimental data.

***The goal** of studying the discipline "Data Analysis" is the formation of students' ability to carry out innovative activities related to the development of software systems for performing data analysis of various structures.*

***The subject of** the "Data Analysis" discipline is methods of software data analysis.*

*The study of the discipline "Data Analysis" strengthens the formation of students **of professional competences (PC)** necessary for solving practical tasks of professional activity:*
***PC08** Ability to apply fundamental and interdisciplinary knowledge to successfully solve software engineering problems.*
***PC15** Ability to apply fundamental and interdisciplinary knowledge to build advanced retrieval algorithms.*
***PC16** Ability to develop software of information retrieval systems.*

*Studying the discipline "Data Analysis" contributes to students' formation of the following **program learning outcomes** (PLO) according to the educational program:*
***PLO01** To analyze, purposefully search for and select for the information and reference resources and knowledge necessary for solving professional tasks, taking into account modern achievements of science and technology.*

***PLO13** To know and apply methods of developing algorithms, designing software, data and knowledge structures.*

PLO25 *To know and to be able to use fundamental mathematical tools to build algorithms and develop modern software.*

PLO31 *To know and be able to apply the principles of building retrieval systems, methods and algorithms for performing various types of information retrieval in them, criteria for evaluating the effectiveness of information retrieval.*

2. Prerequisites and post-requisites of the course (the place of the course in the structural-logical scheme of studies in accordance with educational program)

The successful study of the discipline "Data Analysis" is preceded by the study of the disciplines "Algorithms and data structures" and "Algorithmic support of multimedia and information retrieval systems " of the curriculum of bachelor's training in the specialty 121 Software Engineering.

The theoretical knowledge and practical skills obtained as a result of mastering the discipline "Data Analysis" can be useful for conducting scientific research and completing bachelor's qualification work.

3. Content of the course

The discipline "Data Analysis" involves the study of topics:

Topic 1. Basic provisions of data analysis

Topic 2. Verification of statistical hypotheses

Topic 3. Dispersion analysis

Topic 4. Correlation analysis

Topic 5. Factor analysis

Topic 6. Cluster analysis

Modular control work

Test

4. Educational materials and resources

Basic literature:

1. Methods of data analysis: study guide for students / V.E. Bakhrushin – Zaporizhzhia: KPU, 2011. – 268 p.

Additional literature:

1. Benjamin S. Duran, Patrick L. Odell Cluster Analysis a Survey. – Springer Verlag. – Berlin-Heidelberg New York. - 1974.

Use to master practical skills in the discipline.

2. Jared Dean Big Data Mining, and Machine Learning [Electronic resource], 2014. Access mode: https://www.booksfree.org/wp-content/uploads/2022/06/Big-Data-Data-Mining-and -Machine-Learning-by-Jared-Dean-pdf-free-download-booksfree.org_.pdf

Use to master the theoretical material of the discipline.

3. *Mohammed J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms [Electronic resource], 20 14 . Access mode: http://pzs.dstu.dp.ua/DataMining/bibl/mohammed_j_zaki_wagner_meira_jr_data_mining_and_analysis_fun.pdf*

Use to master practical skills in the discipline.

5. Jiawei Han, Micheline Kamber, Jian Pei Data Mining : Concepts and Techniques . Third Edition [Electronic resource], 201 2 . Access mode: http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data - Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

Use to study the principles of solving nonlinear equations. The materials are freely available on the Internet.

Educational content

5. Methodology of mastering the discipline (educational component)

No.	Type of training session	Description of the training session
Topic 1. Basic provisions of data analysis		
1	*Lecture 1. Classification of data analysis methods*	*Correlation analysis. Analysis of variance. Regression analysis. Covariance analysis. Discriminant analysis. Cluster analysis. Time series analysis*
2	*Lecture 2. Variational statistics*	*Theoretical provisions of variational statistics and construction examples*
3	*Computer workshop 1*	*Task: To develop a program for constructing typical graphs of variational statistics*
Topic 2. Verification of statistical hypotheses		
4	*Lecture 3. Basic provisions of statistical hypotheses*	*Basic concepts. Parametric tests. Non-parametric tests. Determination of empirical data distribution models.*
5	*Lecture 4. Identification of a function*	*Identification of the distribution function of a homogeneous sample. Identification of the distribution function of a heterogeneous sample.*
6	*Computer workshop 2*	*Task: To develop a program for graphical interpretation of empirical data distribution models*
Topic 3. Dispersion analysis		
7	*Lecture 5. One-factor analysis*	*Factorial or between-group variation. Residual or intragroup variation. Kruskel-Wallis rank univariate analysis. Jonkhier-Terpstra criterion*
8	*Lecture 6. Two-factor analysis*	*Analysis of variance by two features. Friedman's ranking criterion. Page's criterion. An example of variance analysis*
9	*Computer workshop 3*	*Task: Develop a program for one-factor and two-factor variance analysis*
10	*Modular control work. Part 1*	
Topic 4. Correlation analysis		

11	Lecture 7. Correlation analysis of quantitative signs	Selective coefficient of determination. Pearson's correlation coefficient. Fechner's correlation coefficient. Covariance matrix
12	Computer workshop 4	Task: Develop a program for calculating correlation coefficients
13	Lecture 8. Correlation analysis of ordinal signs	Rank correlation. Spearman's rank correlation coefficient. Kendall's rank correlation coefficient.
14	Lecture 9. Correlation analysis of nominal signs	Φ - Pearson's coefficient. Root mean square conjugation. Jaccard similarity index
15	Computer workshop 5	Task: To develop a program for rank similarity analysis
16	Lecture 10. Correlation analysis of mixed signs	Gauer coefficient. Biserial correlation coefficient. Biserial correlation coefficient. According to the Kelly-Wood table.
17	Lecture 11. Multiple correlation	Partial correlation coefficient. Multiple correlation coefficient. Canonical correlation analysis. Coefficient of concordance.
18	Computer workshop 6	Task: Develop a program for calculating the coefficient of multiple correlation and concordance
	Topic 5. Factor analysis	
19	Lecture 12. The method of principal components	Factor mapping matrix. Kaiser criterion. Screening criterion.
20	Lecture 13. The method of main factors	Full factorial matrix. Factor mapping. Fundamental theorem of factor analysis
21	Computer workshop 7	Task: Develop a program for factor analysis
	Topic 6. Cluster analysis	
22	Lecture 14. Nearest neighbor method and distant neighbor method	Measures of similarity (dissimilarity). Information statistics. Spearman distance. Kendall Distance. Euclidean distance. Weighted Euclidean distance. Hierarchical and non-hierarchical methods. Nearest and farthest neighbor method
23	Lecture 15. The method of average connection and the method of centers of gravity	Generalized K-distance. Hamming distance. The method of average connection and the method of centers of gravity. Examples of the method of average connection and the method of centers of gravity
24	Computer workshop 8	Task: Develop a program for cluster analysis
25	Lecture 16. The k- means method	k - means method and its modifications. Examples of the algorithm that implements the k - means method
26	Computer workshop 9	Results
27	Modular control work. Part 2	

6. Independent work of a student/graduate student

The discipline "Data Analysis" is based on independent preparation for classroom classes on theoretical and practical topics.

No. z/p	The name of the topic submitted for independent processing	Number of hours	literature

1	*Preparation for lectures*	*16*	*1-5*
2	*Preparation for a computer workshop*	*27*	*1-5*
3	*Preparation for modular control work. Part 1*	*9*	*1-5*
4	*Preparation for modular control work. Part 2*	*9*	*1-5*
5	*Preparation for the test*	*5*	*1-5*

Policy and Assessment

7. Policy of academic discipline (educational component)

Attending classes. *Absence from a classroom session does not involve the calculation of penalty points, since the student's final rating score is formed solely on the basis of the evaluation of study results. At the same time, discussion of the results of the thematic tasks, as well as presentation / public speaking and participation in discussions and additions at seminars will be evaluated during classroom classes. In order to actively participate in the work of the seminar, the student prepares for a specific seminar class in literature as recommended by the teacher. Participation in the work of the seminar also involves the preparation of reports and co-reports within all classes.*

Missed evaluation control measures. *Every student has the right to make up lessons missed for a valid reason (hospital, mobility, etc.) at the expense of independent work. More details at the link: https://kpi.ua/files/n3277.pdf .*

The procedure for contesting the results of assessment control measures. *A student may raise any issue relating to the assessment procedure and expect it to be dealt with in accordance with pre-defined procedures. Students have the right to challenge the results of control measures with arguments, explaining which criteria they disagree with according to the evaluation. Calendar control is carried out in order to improve the quality of students' education and monitor the student's fulfillment of the syllabus requirements.*

Academic integrity. *The policy and principles of academic integrity are defined in Chapter 3 of the Code of Honor of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". More details: https://kpi.ua/code .*

Norms of ethical behavior. *Standards of ethical behavior of students and employees are defined in Chapter 2 of the Code of Honor of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". More details: https://kpi.ua/code .*

Inclusive education. *The acquisition of knowledge and skills in the course of studying the discipline "Research activity in computer engineering" can be accessible to most people with special educational needs, except for students with serious visual impairments that do not allow them to perform tasks with the help of personal computers, laptops and/or other technical means.*

Studying in a foreign language. *In the course of the tasks, students may be recommended to refer to English-language sources. Assigning incentive and penalty points According to the Regulation on the system of evaluation of learning results, the sum of all incentive points cannot exceed 10% of the rating scale.*

All students must attend lectures and practical classes, where you need to actively work on learning the learning material. For objective reasons (for example - illness, international internship), training can take place in an online form individually upon agreement with the head of the course.

Deadlines and Rescheduling Policy:

Works that are submitted late without good reason will be assigned a lower grade. Rearranging modules takes place with the permission of the dean's office if there are good reasons (for example, sick leave).

***Policy on academic integrity** :*

All written works are checked for plagiarism and accepted for defense with correct textual borrowings of no more than 20%. Write-offs during control work are prohibited (including using mobile devices).

8. Types of control and rating system of assessment of learning outcomes

During the semester, students perform 8 computer workshops. The maximum number of points for each computer workshop: 6 points.

Points are awarded for:
- quality of performance of the computer workshop: 0-2 points;
- answer to theoretical questions during the defense of the computer workshop: 0-2 points;
- timely presentation of work for defense: 0-2 points.

Performance evaluation criteria:
2 points – the work is done qualitatively, in full;
1 point - the work is completed in full, but contains minor errors;
0 points – the work is incomplete or contains significant errors.

Answer evaluation criteria:
2 points – the answer is complete, well-argued;
1 point – the answer is generally correct, but has flaws or minor errors;
0 points - there is no answer or the answer is incorrect.

Criteria for evaluating the timeliness of work submission for defense:
2 points – the work is presented for defense no later than the specified deadline;
0 points – the work is submitted for defense later than the specified deadline.

The maximum number of points for performing and defending computer practicals:
6 points × 8 comp. practice = 48 points.

*The assignment for **the modular test** consists of 3 questions - 1 theoretical and 2 practical. The answer to a theoretical question is worth 6 points, and the answer to a practical question is worth 10 points.*

Evaluation criteria for each theoretical test question:
6 points – the answer is correct, complete, well-argued;
5 points – the answer is correct, detailed, but not very well argued;
4 points - in general, the answer is correct, but has shortcomings;
3 points – there are minor errors in the answer;
1-2 points – there are significant errors in the answer;
0 points - there is no answer or the answer is incorrect.

Evaluation criteria for the practical test question:
9-10 points – the answer is correct, the calculations are completed in full;
7-8 points - the answer is correct, but not very well supported by calculations;
5-6 points - in general, the answer is correct, but has flaws;
3-4 points – there are minor errors in the answer;
1-2 points – there are significant errors in the answer;
0 points - there is no answer or the answer is incorrect.

The maximum number of points for a modular control work:
*2 papers * (6 points × 1 theoretical question + 10 points × 2 practical questions) = 52 points.*

The rating scale for the discipline is equal to:

$R_c = R_{com.practice} + R_{MKR} = 48$ points + 52 points = 100 points.

Calendar control: is carried out twice a semester as a monitoring of the current state of fulfillment of the syllabus requirements.
At the first certification (7th week), the student receives "passed" if his current rating is at least 50% of the maximum number of points (20 points) that the student can receive before the first certification.
At the second certification (13th week), the student receives "passed" if his current rating is at least 50% of the maximum number of points (35 points) that the student can receive before the second certification.

Semester control: ***assessment***

Conditions for admission to semester control:
With a semester rating (R_C) of at least 60 points and the enrollment of all computer practical work, the graduate student receives credit "automatically" according to the table (Table of correspondence of rating points to grades on the university scale). Otherwise, he has to complete the credit control work.

Completion and protection of a computer workshop is a necessary condition for admission to the performance of credit control work.

A graduate student can try to improve his grade by writing a graded test, and his semester marks will be canceled ("hard" grading system).

The composition and evaluation criteria of the assessment test:

The test task *consists of 4 questions - 2 theoretical and 2 practical. The answer to each theoretical and practical question is evaluated by 25 points.*

Evaluation criteria for each theoretical test question:
24-25 points – the answer is correct, complete, well-argued;
21-23 points – the answer is correct, detailed, but not very well argued;
17-20 points - in general, the answer is correct, but has flaws;
12-16 points – there are minor errors in the answer;
1-11 points – there are significant errors in the answer;
0 points - there is no answer or the answer is incorrect.

Evaluation criteria for the practical test question:
24-25 points – the answer is correct, the calculations are completed in full;
21-23 points - the answer is correct, but not very well supported by calculations;
17-20 points - in general, the answer is correct, but has flaws;
12-16 points – there are minor errors in the answer;
1-11 points – there are significant errors in the answer;
0 points - there is no answer or the answer is incorrect.

The maximum number of points for a modular control work:
25 points × 2 theoretical questions + 25 points × 2 practical questions = 100 points.

Table of correspondence of rating points to grades on the university scale :

Scores	Grade
100-95	Excellent
94-85	Very good
84-75	Good
74-65	Satisfactory
64-60	Sufficient
Less than 60	Fail
Admission conditions not met	Not Graded

9. Additional information on the discipline (educational component)

The list of questions submitted for semester control will be announced at the last class.

Work program of the academic discipline (syllabus):

Is designed by Ph.D., Assoc. Prof., Onai M.V.

Adopted by Computer Systems Software Department (protocol № 12, 11 February 2026)

Approved by the Methodical commission of the Faculty of Applied Mathematics (protocol № 7, 23 February 2026)