Syllabus VK Intelligent data analysis

Subjects:	Intelligent data analysis
Level of higher education:	first (undergraduate)
Course page in Moodle:	https://dl2022.khadi-kh.com/course/view.php?id=946
The scope of the	4 credits (120 hours)
educational	
component	
Final control form	credit
Consultations:	on schedule
Name of the department:	department of computer technologies and
	mechatronics
Teaching language:	English
Course leader:	Shulyakov Vladyslav Mykolayovych, Ph.D., associate
	professor
Contact phone number:	+38 (067) 737-68-69
Email:	jason07@ukr.net

Brief content of the educational component:

The goal istheoretical and practical training of students to study data processing systems and principles of intelligent data analysis based on Data Mining methods and algorithms.

Subject:creation and research of mathematical and software models of intellectual analysis of data related to the functioning of objects of professional activity.

The main tasks of studying the academic discipline are theoretical and practical training of future specialists on the following issues:

- formation of students' imagination about the application of data storage and organization technologies;
- Data Mining methods and algorithms;
- knowledge discovery processes;
- principles of building data warehouses;
- methods of visual presentation of data.

Prerequisites for studying the educational component:

- "Higher mathematics";
- "Discrete Math";
- "Probability theory and random processes";
- "Theory of algorithms";
- "Numerical Methods";
- "Mathematical methods of operations research";
- "Organization of databases and knowledge".

Competencies acquired by the acquirer:

General competences:

- 1. the ability to communicate in the state language both orally and in writing;
- 2. the ability to search, process and analyze information from various sources;
- 3. ability to abstract thinking, analysis and synthesis;
- 4. the ability to apply knowledge in practical situations.

Special competences:

- 1. the ability to create software for storing, extracting and processing data;
- 2. the ability to accumulate, process and systematize professional knowledge about creating and maintaining software and recognizing the importance of lifelong learning.

Learning outcomes according to the educational program:

- 1. understand, analyze, purposefully search for and choose information and reference resources and knowledge necessary for solving professional tasks, taking into account modern achievements of science and technology;
- 2. know and be able to use methods and means of gathering, formulating and analyzing software requirements;
- 3. know, understand and apply effective approaches to software design;
- 4. know, understand and apply professional standards and other regulatory documents in the field of software engineering;
- 5. know, understand and apply relevant mathematical concepts, methods of domain, system and object-oriented analyzes and mathematical modeling for software development.

Topic No		Number of hours		
	Name of topics (LK, LR, PR, SZ, SR)	ocular	extramural	
1	Lecture #1: Fundamentals of intelligent data analysis.	2		
	Practical lesson 1:Primary exploratory data analysis. Mastering methods of statistical data processing.	4		
	Tasks for independent work: Development of database technology. Comparison of Statistics, Machine Learning and Data Mining.	9		
2	Lecture #2: Regression analysis and non- linear models	2		
	Practical lesson 2: Cluster analysis of data. Mastering the method of building clustering models. Identification of non- obvious regularities in the data.	4		
	Tasks for independent work: Multiple regression analysis. Nonlinear multiple regression model. Non-linear estimation of parameters.	9		
3	Lecture #3: Methods of multivariate intelligence analysis	2		
	Practical lesson 3: Application of classification trees in solving problems of intelligent data analysis. Data preparation and feature selection.	4		
	Tasks for independent work: PAM algorithm. Preliminary dimensionality reduction. Factor analysis. Iterative clustering in SPSS. Raster	9		
4	Lecture #4:Methods of classification	2		

Thematic plan

	Practical lesson 4: Time series	4	
	forecasting. Isolation of periodic		
	components. Time series analysis.		
	Tasks for independent work:	9	
	Methods of support vectors, "nearest	-	
	neighbor", or systems of reasoning based		
	on similar cases. Bayes method. Linear		
	SVM. Object classification in the case of		
	unknown data distributions.		
5	Lecture #5:Forecasting methods	2	
	Practical lesson 5:Correlation-regression	4	
	analysis. Finding the dependence of an		
	effective feature on one or more feature-		
	factors.		
	Tasks for independent work:	9	
	Solution of the forecasting task. Analysis	-	
	of multivariate groupings.		
	Statistical processing of time series and		
	forecasting.		
6	Lecture #6: Methods of finding data	2	
-	patterns		
	Practical lesson 6: Processing of text data	4	
	using Text Miner. Analysis of text data		
	using Text Miner.		
	Taska for independent merke	0	
	I GOZO TAR INAGANANAT WARZ		
	Limits of summer and valiability of the	,	
	Limits of support and reliability of the	7	
	Limits of support and reliability of the associative rule. Methods of finding	7	
	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori	7	
	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm.	2	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining	2	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7:	2 4	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using	2 4	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform.	2 4	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work:	9 2 4 9	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining	9 2 4 9	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools.	9 2 4 9	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual	9 9 2 2	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data	9 2 4 9 2	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8:	9 9 2 4 9 2 4	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed	9 9 2 4 2 4	
7 8	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic	9 9 2 4 9 4	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform.	2 4 9 2 4	
7 8	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work:	9 2 4 9 2 4 4 9	
7 8	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work: Parallel coordinates. Visualization quality.	9 2 4 9 2 4 4 9	
7	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work: Parallel coordinates. Visualization quality. Presentation of spatial characteristics.	9 2 4 9 2 4 4 9 9	
7 8	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work: Parallel coordinates. Visualization quality. Presentation of spatial characteristics. Main trends in visualization.	9 2 4 9 2 4 9 9 9	
7 8 Together	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work: Parallel coordinates. Visualization quality. Presentation of spatial characteristics. Main trends in visualization. Lectures	9 2 4 9 2 4 9 9 9 9 9	
7 8 Together	Limits of support and reliability of the associative rule. Methods of finding associative rules. Varieties of the Apriori algorithm. Lecture #7:OLAP and Data Mining Practical lesson 7: Descriptive and predictive analytics using the Deductor Studio Academic platform. Tasks for independent work: Data Mining process. Data Mining Standards. Data Mining Tools. Lecture #8:Methods of visual presentation of data Practical lesson 8: Reporting and visualization of processed data using the Deductor Studio Academic platform. Tasks for independent work: Parallel coordinates. Visualization quality. Presentation of spatial characteristics. Main trends in visualization. Lectures Practices	9 2 4 9 2 4 9 9 9 9 16 32	

Individual educational and research task: not provided.

Teaching methods:

1) verbal: 1.1 traditional: lectures, explanations, stories, etc.;

1.2 interactive (non-traditional): problem lectures, discussions, etc.;

2) visual: the method of illustrations, the method of demonstrations

3) practical: 3.1 traditional: practical classes, seminars;

3.2 interactive (non-traditional): trainings, "round table", brainstorming method.

Evaluation system and requirements:

The concretization and detailing of the criteria and evaluation system, taking into account the specifics of the educational component, is carried out on the basis of the general criteria specified in STVNZ 7.1-02:2018 "Regulations on the Organization of the Educational Process at the Khnadu" (https://www.khadi.kharkov.ua/fileadmin/P_Ychebotdel/norm_dok/stvnz_7_1_02.pdf) and STVNZ 90.1-01:2021 "Evaluation of learning outcomes of higher education applicants" (https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz-90.1-01_2021.pdf).

Current performance

1 The current success of applicants for the performance of educational types of work in training sessions and for the performance of independent work tasks is evaluated using a four-point rating scale with subsequent transfer to a 100-point scale. During the evaluation of the current performance inare countedall types of work provided by the curriculumprogram

- **1.1** Lectures occupationare evaluated by definition quality implementation of specified tasks
- **1.2** Practical classes are evaluated by the quality of performance of a control or individual task, performance and design of practical work.
- **1.3** Laboratory occupation are evaluated quality
- implementation reports about implementationlaboratory work.
- **1.4** Seminary occupation are evaluated quality implementation individualassignment/abstract.

2 Evaluation of the current success rate of higher education applicants is carried out at each practical session(laboratoryor seminary) on a four-point scale("5", "4", "Z", "2")and are enteredinaccounting journalacademicsuccess

 – "excellent": acquirerflawlessly mastered the theoretical material, demonstrates deepknowledge of relevant topic or academic discipline, main provisions;

– "good": the applicant has mastered the theoretical material well, possesses the main aspects from primary sources and recommended literature, presents it in a reasoned manner; has practical skills, expresses his thoughts on certain problems, but certain inaccuracies and errors are assumed in the logic of the presentation of theoretical content or in the analysis of practical ones;

– "satisfactory": the applicant has basically mastered the theoretical knowledge of the educational topic or discipline, orients himself in primary sources and recommended literature, but answers unconvincingly, confuses concepts, answers additional questions uncertainly, does not have stable knowledge; when answering questions of a practical nature, reveals inaccuracy in knowledge, does not know how to evaluate facts and phenomena, connect them with the future profession;

- "unsatisfactory": the applicant has not mastered the educational material of the topic (discipline), does not know scientific facts, definitions, hardly orients himself in primary sources and recommended literature, lacks scientific thinking, practical skills are not formed.

3 Final scorebycurrent activity is recognized as an arithmetic averagesumpoints for

each lesson, for individual work, current control works according to the formula:

$$K^{current} = \frac{K1 + K2 + \dots + Kn}{n}$$

where $K^{current}$ – final assessment of success based on the results of current control; K1, K2,..., Kn- evaluation of success *n*-th measure of current control;

n- the number of measures of current control.

Grades are converted inpoints according to the calculation scale (table 1).

Table 1– Recalculation of the average grade for the current activity into a multipoint scale

4-point scale	100- point scale	4-point scale	100- point scale	4-point scale	100- point scale	4-point scale	100-point scale
5	100	4.45	89	3.90	78	3.35	67
4.95	99	4.4	88	3.85	77	3.3	66
4.9	98	4.35	87	3.80	76	3.25	65
4.85	97	4.3	86	3.75	75	3.2	64
4.8	96	4.25	85	3.7	74	3.15	63
4.75	95	4.20	84	3.65	73	3.1	62
4.7	94	4.15	83	3.60	72	3.05	61
4.65	93	4.10	82	3.55	71	3	60
4.6	92	4.05	81	3.5	70	from 1.78 to	from 35 to
						2.99	59
						reassembly	
4.55	91	4.00	80	3.45	69	from 0 to 1.77	from 0 to 34
4.5	90	3.95	79	3,4	68	repeated stu	udy

Final assessment

1 The exam is held after studying everyonetopics of the discipline and is composed of students of higher education during the examination session after the end of all classroom classes

2 Applicants of higher education who have completed all types of work prescribed by the curriculum in the discipline are admitted to the exam:

- were present ateveryoneclassroom classes (lectures, seminars, practical);
- completed all missed classes on time;

- scored the minimum number of points for the current academic performance (at least 60 points, which responds on a national scale"3");

If the current performance in the discipline is lower than 60 points, the higher education applicant has the opportunity to increase his current point to the minimum before the beginning of the examination session.

3 Assessment of the knowledge of applicants when taking the exam is carried out on a 100-point scale.

Assessment of the knowledge of applicants through testing is carried out according to the following scale:

- "Excellent": at least 90% of correct answers;
- "Very good": 82% to 89% correct answers;
- "Good": from 74% to 81% of correct answers;
- "Satisfactory": from 67% to 73% of correct answers;
- "Fair enough": 60% to 66% correct answers;
- "Unsatisfactory": less than 60% of correct answers.

4 The final grade for the academic discipline is defined as a weighted average grade that takes into account the overall grade for the current academic performance and the grade for passing the exam.

5 The calculation of the overall final grade for the study of an academic discipline is carried out according to the formula:

$$PK^{ekz} = 0.6 \cdot K^{\text{pumb}} + 0.4 \cdot IS$$

where PC^{ex} - final assessment of success in disciplines, in the form of a final report controls for which there is an exam;

 $K^{current}$ - final assessment of success based on the results of current control (on a 100-point scale);

IS- assessment based on the results of the exam (on a 100-point scale).

0.6 and 0.4

- coefficients of the ratio of points for current success and taking the exam.

6 For performing individual independent work and participating in scientific events, winners are awarded additional points.

6.1 Additional points are added to the total points scoredcakeof higher education for the current educational activity (for disciplines for which the final form of control is a credit), or to the final assessmentwithdisciplines, the final form of control for which is an exam.

6.2 The number of additional points awarded for different types of individual tasks depends on their volume and importance:

 prizes in the discipline at the international / all-Ukrainian competition of scientific student works - 20 points;

- prize places in the discipline at the All-Ukrainian Olympiads - 20 points;

- participation in the international / all-Ukrainian competition of scientific student works
- 15 points

 participation in international / all-Ukrainian scientific conferences of students and young scientists - 12 points;

- participation in all-Ukrainian Olympiads in the discipline 10 points
- participation in olympiads and scientific conferences of the Khnadu in the discipline 5 points;

- performance of individual scientific and research (educational and research) tasks of increased complexity - 5 points.

6.3 The number of additional points cannot exceed 20 points.

7 General final rating by study educational disciplines not maybeexceed 100 points.

The overall final grade for the study of the academic discipline is determined according to the scale given in Table 2.

Table 2– The scale for evaluating the knowledge of the students based on the results of the final control of the academic discipline

Score	Asse	ssment	Evaluation according to the ECTS scale		
IN points	by natio	by national scale		Criteria	
points	examinat	credit			
	ion				
90-100	Perfectly	Enrolled	A	The theoretical content of the course has been mastered in its entirety, without gaps, the necessary practical skills for working with the mastered material have been formed, all the educational tasks provided for in the training program have been completed, the quality of their performance has been assessed by the number of points close to to the maximum	
80-89	Okay	Okay Iled		The theoretical content of the course has been mastered in its entirety, without gaps, the necessary practical skills for working with the mastered material have mainly been formed, all educational tasks provided for in the training program have been completed, the quality of most of them has been evaluated with a number of points close to the maximum	
75-79		Enroll	WITH	The theoretical content of the course has been mastered completely, without gaps, some practical skills of working with the mastered material have not been sufficiently developed, all the educational tasks provided for by the training program have been completed, the quality of none of them has been evaluated with a minimum number of points, some types tasks were completed with errors	
67-74	ctorily	ctorily olled		The theoretical content of the course is partially mastered, but the gaps are not of a significant nature, the necessary practical skills for working with the mastered material are basically formed, most of the educational tasks provided by the training program have been completed, some of the completed tasks, may contain errors	
60–66	Satisfac	E	The theoretical content of the course has been partially mastered, some practical work skills have not been formed, many of the educational tasks provided by the training program have not been completed, or the quality of some of them has been assessed with a number of points close to minimal		

35–59	Unsatisfactorily	FX		The theoretical content of the course has been partially mastered, the necessary practical work skills have not been formed, most of the prescribed training programs of educational tasks have not been completed, or the quality of their implementation has been assessed with a number of points close to the minimum; with additional independent work on the course material, it is possible to improve the quality of the performance of educational tasks (with the possibility of repeated drafting)
0–34	Unacceptable	Not counted	F	The theoretical content of the course has not been mastered, the necessary practical work skills have not been formed, all completed educational tasks contain gross errors, additional independent work on the course material will not lead to any significant improvement in the quality of the performance of educational tasks (with a mandatory repeat course)

Course policy:

- the course involves working in a team, the environment in the classroom is friendly, creative, open to constructive criticism;

- mastering the discipline involves mandatory attendance at lectures and practical classes, as well as independent work;

- independent work involves the study of individual topics of the academic discipline, which are presented in accordance with the program for independent study, or were considered briefly;

- all tasks provided by the program must be completed within the set time;

- if the student of higher education is absent from classes for a valid reason, he presents the completed tasks during independent preparation and consultation of the teacher;

- course work must be protected no later than a week before the beginning of the examination session;

- while studying the course, students of higher education must adhere to the rules of academic integrity set forth in the following documents: "Rules of Academic Integrity

participants educational process LOOKING

 FOR"(<u>https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_67_01_dobro</u>

 ch_1.p_df), "Academic integrity. Verification of academic, scientific and qualification texts on

plagiarism"

(https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_85_1_01.pdf),

"Moral and ethical code of participants in the educational process of the National Academy of Sciences (https://www.khadi.kharkov.ua/fileadmin/P_Standart/pologeniya/stvnz_67_01_MEK_1.

<u>pdf</u>).

- in case of detection of plagiarism, the applicant receives 0 points for the task and must repeat the tasks provided for in the syllabus;

writing off during tests and exams is prohibited (including using mobile devices).
 Mobile devices are allowed to be used only during online testing.

Recommended Books:

1. Ian H. Witten Data Mining: Practical Machine Learning Tools and Techniques (Paperback). – 1999, 525 p.

2. Michael JA Berry Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (Paperback). – 1997, 643 p.

3. Foster Provost Data Science for Business: What you need to know about data mining and data-analytic thinking (Paperback). – 2013, 414 p.

Additional sources:

1. Galit Shmueli, Nitin R. Patel , Peter C. Bruce Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with Xlminer. - 2006, 428 p.

2. Wes McKinney Python for Data Analysis. – 2011, 400 p.

3. Anand Rajaraman, Jeffrey D. Ullman Mining of Massive Datasets. – 2011, 326 p.

4. Distance course-resource "Intellectual data analysis" [Electronic resource]. – Access mode: https://dl2022.khadi-kh.com/course/view.php?id=946.

Developer(s) the syllabus of the academic discipline Ph.D. technical of Science, Assoc.

_ Vladyslav Shulyakov

signature

Reeree

Head of the KTM department dr. technical Sciences, Prof.

Oleg NIKONOV

signature