**Educational Establishment
"Grodno State Agrarian University"**

**APPROVED
rector
_____ V.K.Pestis
"___" _____ 2019
Registration № UD ____**

TECHNOLOGIES OF DATA INTELLECTUAL ANALYSIS

**The curriculum of higher education institutions for undergraduates
1-25 80 01 "Economics" (profiling - "Agricultural Economics")**

2019 y.

COMPILED BY:

E. A. SUCHANOVA, lecturer of the department of computer science and economic and mathematical modeling in the agricultural sector
educational establishment "Grodno state agrarian university"

RECOMMENDED FOR APPROVAL:

Department of computer science and economics and mathematical modeling in the agricultural sector
(protocol № 15 on 24.05.2019);
Methodical council of educational establishment "Grodno state agrarian university"
(protocol № _ on __.__.____)
Responsible for editorial   E. A. Suchanova
Responsible for release   T.V. Snopko

EXPLANATORY NOTE

The goals and objectives of the discipline

The purpose of the study of the discipline is the formation of knowledge and skills among undergraduates in the use of technologies of high-performance intellectual analytical information processing, aimed at quickly extracting valuable expert knowledge from significant arrays of accumulated and incoming data, while supporting effective management activities.

Objectives of the discipline:

− give basic theoretical knowledge about the basic concepts and methods of data mining;

− develop skills and abilities for using software based on data mining technology to solve practical problems.

Place of discipline in the system of training a specialist with higher education

The discipline is included in the "Component of higher education institution" section of the curriculum for undergraduates in the specialty 1-25 80 01 "Economics" (specialization "Agricultural Economics").

The mastery of the discipline is based on the competencies acquired earlier by undergraduates in the study of the disciplines "Higher Mathematics", "Computer Information Technologies", "Statistics".

Requirements for the level of development of the content of the academic discipline

As a result of studying the discipline, the undergraduate must consolidate and develop the following academic (AC) and social-personal (SPC) competencies:

AC-1. To be able to apply basic scientific and theoretical knowledge to solve theoretical and practical professional problems.

AC-2. Be able to do comparative analysis.

AC-3. Be able to work independently.

AC-4. Have an interdisciplinary approach to solving problems.

AC-5. Have skills when working with computer equipment.

SPC-1. Have the ability to interpersonal communications.

SPC-2. Be able to work in a team.

As a result of studying the discipline, the undergraduate must possess the following professional competencies (PC):

PC-1. Make decisions on choosing the optimal informational structure of production.

PC-2. Conduct organizational and management survey of production.

PC-3. Choose the most effective forms of adaptation of the enterprise to environmental conditions.

PC-4. Use modern information technologies and applied data processing programs to substantiate management decisions.

PC-5. Make effective management decisions in the economic sphere.

PC-6. Be able to analyze the organization's activities in international markets and develop strategies for their conquest.

PC-7. To evaluate the results of economic activities of organizations.

PC-8. Work with scientific, technical and patent literature.

To acquire professional competencies PC-1 - PC-8 as a result of studying the discipline, the undergraduate must know:

- tasks of decision support systems;
- tasks of Data Mining;
- practical application of Data Mining;
- Data Mining models;
- Data Mining methods;
- Knowledge discovery process;
- Data Mining tools;
- Scopes of Data Mining.

To be able and able to:

- qualify the tasks of Data Mining;
- apply Data Mining tools.

Total hours and class hours

100 hours are maximum allocated for mastering the discipline, including 42 hours of classroom hours.

Higher Education Form

When studying the discipline, such forms of training are provided: full-time and part-time.

Classroom time distribution by occupation, course and semester

Discipline is studied for full-time study in the 1st semester, while the class time is divided into 18 hours of lecture and 24 hours of laboratory classes.

Forms of current certification in academic discipline

Assessment of academic achievements of a graduate student is an credit. To assess the educational achievements of undergraduates, criteria are used, approved by the Ministry of Education of the Republic of Belarus.

Evaluation of intermediate academic achievements of undergraduates is carried out in accordance with the selected grade scale.

To assess the achievements of undergraduates, the following diagnostic tools are used (competencies are checked):

- presentation by a graduate student at a conference on a prepared abstract (AC-1, AC-3, AC-4, AC-5, SPC-1, SPC-2, PC-7);

- conducting ongoing quizzes on selected topics (PC-1 - PC-8);

- protection of individual tasks performed in laboratory studies (AC-3, AC-5, SPC-1, PC-1– PC-8);

- protection of individual tasks performed within the framework of controlled independent work (AC-1, AC-3, AC-5, SPC-1, PC-1 - PC-8);

- passing the credit in the discipline (AC-1 - AC-5, PC-1 - PC-8).

CONTENT OF EDUCATIONAL MATERIAL

Theme 1. Decision support systems

Tasks of decision support systems. Databases are the basis of DSS. Data Warehousing. OLAP systems

Theme 2. Basic concepts of data mining

Data Mining Tasks. Practical application of Data Mining. Data Mining Models. Data Mining Methods. The process of discovering knowledge. Data Mining Tools.

Theme 3. The main methods of data mining

3.1 Classification and regression

Formulation of the problem. Presentation of results: *Classification rules. Decision trees. Mathematical functions.*

Methods for constructing classification rules *Algorithm for constructing 1-rules. Naive Bayes Method*

Methods for constructing decision trees: *divide and conquer technique. Coverage algorithm.*

Methods for constructing mathematical functions *General view. Linear methods. Least square method. Nonlinear Methods Support Vector Machines (SVM). Regularization Networks. Regularization Networks. Sampling and Rare Grids.*

Time Series Forecasting. Setting goals. *Time Series Forecasting Methods.*

3.2 Search for associative rules

Formulation of the problem. *Formal statement of the problem. Sequential analysis. Varieties of the task of finding associative rules.*

Presentation of the results.

Algorithms *Apriori Algorithm. Varieties of the Apriori algorithm.*

3.3 Clustering

Clusterization problem statement. Formal statement of the problem. Clustering is based on the algorithms used in the algorithms.

Presentation of the results.

Basic clustering algorithms. *Classification of algorithms. Hierarchical Algorithms. Non-hierarchical algorithms.*

Adaptive clustering methods. *Choosing the best solution and the quality of clustering. Use of formal quality criteria in adaptive clustering. An example of adaptive clustering.*

3.4 Neural networks

Application area. Elements of neural networks. The architecture of neural networks. Neural network training. Models of neural networks. Maps of Kohonen. Software for working with neural networks.

3.5 Visual Data Analysis - Visual Mining

Perform visual analysis of data. Characteristics of data visualization tools. Visualization methods.

Theme 4. Data mining tools

Data mining tools and their classification. The use of computer technology to solve the problems of Data Mining.

EDUCATIONAL-METHODICAL MAP

Section number, topics, classes	Section title, topics, classes; list of issues under study	Total hours	The number of class hours				The number of hours devoted to independent work of students (including hours allocated for the implementation of the course work / project	Financial support classes (visual, manuals, etc.)	Literature	The forms of knowledge control
			lectures	practical (seminar) classes	laboratory classes	Independent work of students (CER)				
1	2	3	4	5	6	7	8	9	10	11
Theme 1.	**Decision support systems** Tasks of decision support systems. Databases are the basis of DSS. Data Warehousing. OLAP systems	10	2			8		EEMC and online tutorials featured in the local network of the library, testing through distance learning site	[1]	Testing
Theme 2.	**Basic concepts of data mining** Data Mining Tasks. Practical application of Data Mining. Data Mining Models. Data Mining Methods. The process of discovering knowledge. Data Mining Tools.	12	2			10		EEMC and online tutorials featured in the local network of the library, testing through distance learning site	[1]	Testing
Theme 3.	**The main methods of data mining**	44	14			30		EEMC and online tutorials featured in the local network of the library, testing through distance learning site		Testing
3.1	**Classification and regression**		4			6			[1, 8]	
3.1.1	Formulation of the problem. Presentation of results: *Classification rules. Decision trees. Mathematical functions.* Methods for constructing classification rules *Algorithm for constructing 1-rules. Naive Bayes Method* Methods for constructing decision trees: *divide and conquer technique. Coverage algorithm.*		2							
3.1.2	Methods for constructing mathematical functions *General view. Linear methods. Least square method. Nonlinear Methods Support Vector Machines (SVM). Regularization Networks. Regularization Networks. Sampling and Rare Grids.* Time Series Forecasting. *Setting goals. Time Series Forecasting Methods.*		2							

7

1	2	3	4	5	6	7	8	9	10	11
3.2.	**Search for associative rules** Formulation of the problem. *Formal statement of the problem. Sequential analysis. Varieties of the task of finding associative rules.* Presentation of the results. Algorithms *Apriori Algorithm. Varieties of the Apriori algorithm.*		2			6				
3.3	**Clustering**		4			6			[1. 2]	
3.3.1	Clusterization problem statement. Formal statement of the problem. Clustering is based on the algorithms used in the algorithms. Presentation of the results. Basic clustering algorithms. *Classification of algorithms. Hierarchical Algorithms. Non-hierarchical algorithms.*		2							
3.3.2	Adaptive clustering methods. *Choosing the best solution and the quality of clustering. Use of formal quality criteria in adaptive clustering. An example of adaptive clustering.*		2							
3 4.	**Neural networks** Application area. Elements of neural networks. The architecture of neural networks. Neural network training. Models of neural networks. Maps of Kohonen. Software for working with neural networks.		2			6			[1, 3]	
3.5	**Visual Data Analysis - Visual Mining** Perform visual analysis of data. Characteristics of data visualization tools. Visualization methods.		2			6			[1]	
Theme 4.	**Data mining tools** Data mining tools and their classification. The use of computer technology to solve the problems of Data Mining.	**34**			**24**	**10**		EEMC and online tutorials featured in the local network of the library, testing through distance learning site	[1,2,3,4-8]	Testing
4.1	Classification and Regression				8	2				
4.1.1	*Decision Tree Method Algorithms*				2					lab work report
4.1.2	*Support Vector Method*				2					lab work report
4.1.3	*Nearest Neighbor*				2					lab work report
4.1.4	*Bayesian classification*				2					lab work report
1	2	3	4	5	6	7	8	9	10	11
4.2	Estimates and Forecasting				4	2				

4.2.1	*Forecasting and Time Series*				2				lab work report
4.2.2	*Forecasting and visualization*				2				lab work report
4.3	Associative rule search				4	2			lab work report
4.4	Clustering and Segmentation				4	2			
4.4.1	*Hierarchical methods*				2				lab work report
4.4.2	*Iterative Methods*				2				lab work report
4.5	Neural networks				4	2			lab work report
	Total	100	18		24	58			

9

INFORMATION AND METHODOLOGICAL PART

Monitoring the academic achievements of undergraduates

For current monitoring of academic achievements of undergraduates, tests, multilevel control tasks, oral questioning during classes and other diagnostic tools are used. The final assessment of academic achievement is carried out on the exam. For this, criteria approved by the Ministry of Education of the Republic of Belarus are used.

Sample list of questions for the final control

1. Tell us about the tasks of decision support systems.
2. The database as the basis of decision support systems.
3. Tell us about the architecture of decision support systems.
4. Tell us about the concept of data warehouse.
5. Tell us about the organization of the data warehouse.
6. Tell us about the multidimensional data model.
7. Define the OLAP system, basic concepts.
8. Tell us about conceptual multi-dimensional data presentation.
9. Tell us about the architecture of OLAP-systems.
10. Tell us about the basic concepts of data analysis.
11. Tell us about the challenges of data analysis and areas of practical application.
12. List the main stages of building models.
13. Tell us about the methodology of data analysis.
14. List the tools for database mining.
15. Tell us about the identification of patterns and forecasting by neural network and other modern methods.
16. Tell us about the task of classification and regression. Formulate its formulation, tell about the decision algorithms, presentation of the results.
17. Formulate the problem of searching for associative rules and its formulation. Tell us about decision algorithms, presentation of results.
18. Tell us about the task of clustering. Formulate the statement of the problem, tell about the solution algorithms, presentation of the results.
19. Tell us about performing visual data analysis and visualization methods.

Recommended reading list

Main literature:

1. B. Liu, Web Data Mining, Springer 2006.
2. D. J. Hand, H. Mannila, and P. Smyth, Principles of Data Mining, MIT Press, 2001

3. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann, 2nd ed. 2005

4. J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd ed., 2011

5. P.-N. Tan, M. Steinbach and V. Kumar, Introduction to Data Mining, Wiley, 2005

6. R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, 2ed., Wiley-Interscience, 2000

7. S. Chakrabarti. Mining the Web: Statistical Analysis of Hypertex and Semi-Structured Data. Morgan Kaufmann, 2002

8. S. M. Weiss and N. Indurkhya, Predictive Data Mining, Morgan Kaufmann, 1998

9. T. Dasu and T. Johnson. Exploratory Data Mining and Data Cleaning. John Wiley & Sons, 2003

10. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer-Verlag, 2009

11. U. Fayyad, G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann, 2001

Additional literature:

1. D. P. Ballou and G. K. Tayi. Enhancing data quality in data warehouse environments. Communications of ACM, 42:73-78, 1999

2. D. Pyle. Data Preparation for Data Mining. Morgan Kaufmann, 1999

3. E. Rahm and H. H. Do. Data Cleaning: Problems and Current Approaches. *IEEE Bulletin of the Technical Committee on Data Engineering. Vol.23, No.4*

4. H.V. Jagadish et al., Special Issue on Data Reduction Techniques. Bulletin of the Technical Committee on Data Engineering, 20(4), December 1997

5. R. Wang, V. Storey, and C. Firth. A framework for analysis of data quality research. IEEE Trans. Knowledge and Data Engineering, 7:623-640, 1995

6. T. Dasu and T. Johnson. Exploratory Data Mining and Data Cleaning. John Wiley & Sons, 2003

7. T. Dasu, T. Johnson, S. Muthukrishnan, V. Shkapenyuk. Mining Database Structure; Or, How to Build a Data Quality Browser. SIGMOD'02.

8. T. Redman. Data Quality: Management and Technology. Bantam Books, 1992

9. V. Raman and J. Hellerstein. Potters Wheel: An Interactive Framework for Data Cleaning and Transformation, VLDB'2001

10. Y. Wand and R. Wang. Anchoring data quality dimensions ontological foundations. Communications of ACM, 39:86-95, 1996

PROTOCOL AGREEMENT CURRICULUM SVR

The name of discipline, which requires approval	Name of department	Proposals for changes in the content of the training on the subject matter under study program	Action taken by the Department, develop training programs (With date and protocol number)
Information support production	Department of Informatics and EMM in AIC	No offers	Considered at the meeting of the department, to Protocol N ° 15 on May 24, 2019
Information marketing	Department of Management, Marketing and Law	No offers	Considered at the meeting of the department, to Protocol N ° 15 on May 24, 2019

ADDITIONS AND CHANGES TO EDUCATIONAL PROGRAM SVR
on ____ / _____ school year

№	Additions and changes	Base

The curriculum is reviewed and approved at a meeting of the department
Informatics and EMM in the AIC (protocol № __ from _____ __ 20__)

(the department name)

Head of Department

Candidate of Physics and Mathematics Sciences, Associate Professor _____   T.N.Izosimova

(academic degree, academic title)   (signature)   (Initials and Family names)

APPROVED

Dean of the Faculty

Candidate of Economics. Sciences, Associate Professor _____   A.V. Gribov

(academic degree, academic title)   (signature)   (Initials and Family names)