

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ /
BELARUSIAN STATE UNIVERSITY**

УТВЕРЖДАЮ / APPROVED

Ректор Белорусского

государственного университета/

Rector of Belarusian State University

А.Д.Король /Andrei D.Karol

23.12.2024

Регистрационный

2143/м

№ / Registration №



**ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ/
INTELLIGENT DATA ANALYSIS**

**Учебная программа учреждения образования по учебной дисциплине для
специальности:**

The program of the educational institution of the discipline for the speciality:

7-06-0533-05 Прикладная математика и информатика /

7-06-0533-05 Applied Mathematics and Computer Science

Профилизация / Profilization:

Компьютерный анализ данных / Computer Data Analysis

2024 г.

Учебная программа составлена на основе ОСВО 7-06-0533-05-2023 и учебного плана № М53а-5.3-115/уч. от 11.04.2023.

СОСТАВИТЕЛЬ:

П.А. Пашук, старший преподаватель кафедры теории вероятностей и математической статистики факультета прикладной математики и информатики Белорусского государственного университета, магистр прикладной математики и информационных технологий.

РЕЦЕНЗЕНТЫ:

М.С. Абрамович, заведующий НИЛ статистического анализа и моделирования учреждения Белорусского государственного университета «НИИ прикладных проблем математики и информатики», кандидат физико-математических наук, доцент;

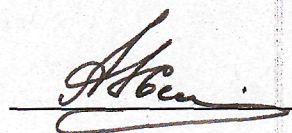
А.И. Кишкар, ведущий инженер-программист отдела информационных систем управления бизнес-приложений департамента производства ЗАО «Международный деловой альянс», магистр прикладной математики и информационных технологий.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики БГУ (протокол № 5 от 19.11.2024);

Научно-методическим советом БГУ (протокол № 5 от 19.12.2024)

Заведующий кафедрой теории вероятностей и математической статистики



А.Ю.Харин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины «Intelligent data analysis» – изучение современных методов и алгоритмов интеллектуального анализа данных и формирование навыков решения практических задач с использованием современного программного обеспечения.

Задачи учебной дисциплины:

- 1) изучение основных подходов и методов интеллектуального анализа данных;
- 2) знакомство студентов с применением методов интеллектуального анализа данных, а также их преимуществами и недостатками;
- 3) формирование практических навыков решения прикладных задач с использованием современного программного обеспечения.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к модулю «Специальные методы анализа в прикладных задачах» компонента учреждения высшего образования.

Учебная программа составлена с учетом межпредметных **связей** и программ по дисциплинам: «Multivariate statistical analysis», «Methods and algorithms of machine learning», «Neural networks in machine learning», «Methods for statistical analysis of complex data», «Data mining methods», «Internet data analysis».

Требования к компетенциям

Освоение учебной дисциплины «Intelligent data analysis» должно обеспечить формирование следующих компетенций:

универсальные компетенции (UC):

UC. To be able to apply scientific cognition methods in research activity, to generate and realize innovative ideas;

специальные компетенции (SC):

SC. To use knowledge of probability models and methods for intelligent data analysis.

В результате освоения учебной дисциплины студент должен:

знать:

- основные методы интеллектуального анализа данных;
- особенности методов и алгоритмов анализа данных;

уметь:

- использовать методы и алгоритмы интеллектуального анализа данных;
- выбирать оптимальный метод интеллектуального анализа данных при решении задачи;
- интерпретировать полученные результаты

иметь навык:

- работы с основными методами анализа данных;
- компьютерной реализации основных методов;

– решения прикладных задач с использованием современного программного обеспечения;

Структура учебной дисциплины

Дисциплина изучается в 2 семестре. В соответствии с учебным планом всего на изучение учебной дисциплины «Intelligent data analysis» отведено для очной формы получения высшего образования – 120 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма промежуточной аттестации – зачет.

EXPLANATORY NOTE

Aim and tasks of the discipline

Aim of the discipline «Intelligent data analysis» – studying modern methods and algorithms for intelligent data analysis and developing skills in solving practical problems using modern software.

Tasks of the discipline:

- 1) study of basic approaches and methods of intelligent data analysis;
- 2) familiarization of students with the use of intelligent data analysis methods, as well as their advantages and disadvantages;
- 3) developing practical skills in solving applied problems using modern software.

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

The academic discipline is part of the module «Special methods for analysis in applied problems» of higher education institution component.

The curriculum is designed taking into account interdisciplinary connections and programs in disciplines: «Multivariate statistical analysis», «Methods and algorithms of machine learning», «Neural networks in machine learning», «Methods for statistical analysis of complex data», «Data mining methods», «Internet data analysis».

Requirements for competences

Mastering of the academic discipline «Intelligent data analysis» should provide the formation of the following universal and advanced professional competences:

universal competencies (UC):

UC. To be able to apply scientific cognition methods in research activity, to generate and realize innovative ideas;

special competencies (SC):

SC. To use knowledge of probability models and methods for intelligent data analysis.

As a result of mastering the academic discipline, the student is expected to:

know:

- basic methods of intelligent data analysis;
- features of methods and algorithms of intelligent data analysis;

be able to:

- use methods and algorithms of intelligent data analysis;
- select the optimal method of intelligent data analysis when solving a problem;

have skills in:

- working with basic methods of intelligent data analysis;
- computer implementation of basic methods;
- solving applied problems using modern software;

Structure of the academic discipline

The discipline is studied in the 2 semester. In total for the study of the discipline «Intelligent data analysis» is allocated for full-time higher education – 120 hours, including 40 in-class hours, of them: lectures – 20 часов, laboratory classes – 20 hours.

The labour intensity of the discipline is 3 credit units.

Form of certification – end-of-term test.

CONTENT OF THE STUDY MATERIAL

Section 1. Introduction to Intelligent data analysis

Topic 1.1. Introduction to Intelligent data analysis.

Methods and stages of Intelligent data analysis. Tasks and areas of application of Intelligent data analysis. The process of knowledge discovery. The difference between Intelligent data analysis and other methods of data analysis.

Section 2. Methods for searching for association rules

Topic 2.1. Types of association rules.

Shopping cart analysis task. Information table. Association rule. Support and reliability of association rule. Types of association rules.

Topic 2.2. Binary association rules.

Search for binary association rules – general statement of the problem. Support and reliability of the measure of association rules. Strong association rules. Other measures of association rule evaluation. Frequent sets. Naive algorithm for finding frequent sets. Basic algorithms for searching association rules. Apriori algorithm. Monotonicity property of the support measure.

Topic 2.3. Multidimensional and multilevel association rules.

Dimensionality of the data under study. One-dimensional and multidimensional association rules. Degree of abstraction of the data under study. Single-level and multilevel association rules. Generation of candidate sets of frequent sets of different dimensions. Varieties of the Apriori algorithm. The idea of the FP Growth algorithm.

Section 3. Search for typical sequences

Topic 3.1. Sequence templates.

Sequence pattern detection problem. Sequence expressions, sequence length and size, maximum sequence. Finding frequent episodes in a sequence.

Topic 3.2. Investigation of typical sequences.

The basic GSP algorithm consists of the following steps: sorting, finding frequent events, finding all frequent sequences and maximal sequences. Study of typical sequences. Prefix Span algorithm. Application areas.

Section 4. Cluster analysis

Topic 4.1. Basic concepts of cluster analysis.

Data types in cluster analysis. Real binary data. Dependency measures. Cosine dependence measure.

Topic 4.2. Special methods of cluster analysis.

Iterative and hierarchical methods of cluster analysis. Graphical representation of clustering results. DBSCAN algorithm. Kohonen self-organizing maps and their use in cluster analysis.

Section 5. Analysis of text documents and web pages

Topic 5.1. Analysis of text documents and web pages.

Text mining. Search for text documents based on a query and based on an existing document. Classification of text documents. Grouping of text documents and citation analysis on the web. Document search evaluation measures. TFM matrix. Text sentiment analysis, problem statement. Methods based on rules and dictionaries. Method based on graph-theoretical models. Quality assessment.

Topic 5.2. Analysis of web pages.

Web mining. Link ranking method PageRank. Statement of the problem of ranking links. Algorithm PageRank based on the power method. Text stemming.

TEACHING AND METHODOLOGICAL MAP OF THE DISCIPLINE
Full-time form of higher education with the use of distance learning technologies (DLT)

Title of section, topic	Title of section, topic	In-class hours					Independent work	Form of control
		Lectures	Practical classes	Seminar classes	Laboratory classes	Other		
1	Introduction to Intelligent data analysis	2			2			
1.1	Introduction to Intelligent data analysis	2			2			Oral test
2	Methods for searching for association rules	6			6			
2.1	Types of association rules	2			2			Individual tasks
2.2	Binary association rules	2			2			Control work
2.3	Multidimensional and multilevel association rules	2			2			Report on the performance of laboratory task
3	Search for typical sequences	4			4			
3.1	Sequence templates	2			2			Oral test, individual tasks
3.2	Investigation of typical sequences	2			2			Test, report on the performance of laboratory task
4	Cluster analysis	4			4			
4.1	Basic concepts of cluster analysis	1			1			Individual tasks
4.2	Special methods of cluster analysis	3			3			Control work, report on the performance of laboratory task
5	Analysis of text documents and web pages	4			4			
5.1	Analysis of text documents	2			2			Report on the performance of

								laboratory task
5.2	Analysis of web pages	2			2			Test, individual tasks
	TOTAL	20			20			

INFORMATION AND METHODOLOGICAL PART

List of basic literature

1. Труш, Н. Н. Введение в компьютерный и интеллектуальный анализ данных: учеб. материалы / Н.Н. Труш. - Минск: БГУ, 2022 - 69 с. - <https://elib.bsu.by/handle/123456789/277034>.
2. Якимов, А. И. Интеллектуальный анализ данных для имитационного моделирования производственных систем / А. И. Якимов, Е. А. Якимов, Е. М. Борчик ; М-во образования Республики Беларусь, М-во науки и высшего образования РФ, МОУ ВО "Белорусско-Российский ун-т". - Могилев : Белорусско-Российский ун-т, 2021. - 183 с.
3. Дюк, В. А. Логический анализ данных : учебное пособие / В. А. Дюк. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2020. - 77 с. - <https://e.lanbook.com/book/126935>

List of additional literature

1. Джеймс, Г. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С.Э. Мастицкого / Г. Джеймс, Д. Уиттон, Т. Хасти, Р. Тибширани. – М.: ДМК Пресс, 2016. – 450 с.
2. Буяльская, Ю.В. Введение в компьютерный и интеллектуальный анализ данных: метод. указания / Ю.В. Буяльская, В.В. Казаченок – Минск: БГУ, 2016. – 46 с.
3. Степанов, Р.Г. Технология DATA MINING: Интеллектуальный анализ данных / Р.Г. Степанов – Казань, Издательство Казанского госуниверситета, 2008 – 58 с.
4. Мусаев, А.А. Интеллектуальный анализ данных: учебное пособие / А.А. Мусаев. – СПб.: СПбГТИ(ТУ), 2018. – 176 с.
5. Рафалович, В. Data mining, или Интеллектуальный анализ данных для занятых. Практический курс / В. Рафалович – Москва: Литагент И-Трейд, 2014. – 96 с.
6. Петрунин, Ю.Ю. Информационные технологии анализа данных. Data analysis. Изд.4 / Ю.Ю. Петрунин – М.: Издательство МГУ, 2023. – 296 с.
7. Зарова, Е.В. Методы Data mining в обработке и анализе статистических данных (решения в R) / Е.В. Зарова – М.: Инфра-М, 2021. – 232 с.
8. Rutkowski L., Jaworski M., Dudo P. Stream Data Mining: Algorithms and Their Probabilistik Properties. – Springer. 2020. – 340 p.
9. Ghavami P. Big Data Analitics Methods. Analitics Techniques in Data Mining. Deep Learning and Natural Language Processing. – De Gruyter Pub., 2019. – 254 p.
10. Rutkowski L., Jaworski M., Dudo P. Stream Data Mining: Algorithms and Their Probabilistic Properties. Springer, 2020. 340 p.

11. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAD. СПб.: БХВ Петербург, 2007. – 384 с.

12. Han J., Kamber M., Morgan Kaufman. Data Mining: concepts and techniques. – Morgan Kaufman Pub., 2012. – p.324.

13. Hand J., Mannila H., Smyth P. Principles of Data Mining. – MIT Press, 2001. – p.261.

List of recommended diagnostic tools and methodology for final mark formation

The object of diagnostics of students' competences is the knowledge and skills acquired as a result of studying the academic discipline. Identification of students' learning achievements is carried out by means of current and interim certification.

The following means of current certification can be used to diagnose competences: control work, tests, report on the performance of laboratory task, individual tasks, oral test.

The form of interim certification in the discipline “Intelligent data analysis” in accordance with the curriculum is end-of-term test.

A rating system of the student knowledge is used for the final mark formation, which makes it possible to trace and evaluate the dynamics within the process of achieving learning objectives. The rating system stipulates the use of weighting coefficients for current and interim certification of students in the academic discipline.

The final mark formation in the course of control measures for current certification (approximate weighting coefficients determining the contribution of current certification to the mark for passing interim certification) includes:

- performance of control work – 20 %;
- performance of test – 20 %;
- performance of individual tasks – 20 %;
- report on the performance of laboratory task – 40 %.

The final mark for the discipline is calculated on the basis of the mark of current certification (rating system of knowledge) — 60 % and end-of-term test mark — 40 %.

Approximate list of laboratory classes

Class № 1. Introduction to Intelligent data analysis.

Class № 2. Types of association rules.

Class № 3. Binary association rules.

Class № 4. Multidimensional and multilevel association rules.

Class № 5. Sequence templates.

Class № 6. Investigation of typical sequences.

Class № 7. Basic concepts of cluster analysis.

Class № 8. Special methods of cluster analysis.

Class № 9. Analysis of text documents.

Class № 10. Analysis of web pages.

Description of innovative approaches and methods for teaching the discipline

When organizing the educational process, a practice-based approach is used, which entails the following:

- mastering the educational content through solving practical tasks;
- acquiring skills for effective performance in various types of professional activities;
- orientation towards idea generation, implementation of students' group projects, development of business culture;
- use of evaluation procedures, assessment methods, indicating the formation of professional competences.

Methodological recommendations for the organization of independent work

Independent work for the purpose of studying the material of the academic discipline involves working with recommended educational literature and Internet resources. Theoretical information is consolidated by completing laboratory assignments, during which one should be guided by the methodological developments posted in the electronic library of the university and on the educational portal. Additional assignments (tests, assignments for independent completion) may also be offered for self-assessment and deeper assimilation of the material received.

Sample list of questions for the end-of-term test

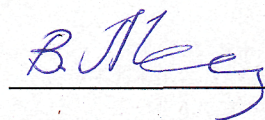
1. Tasks and areas of application of Intelligent data analysis. The process of knowledge discovery.
2. The difference between Intelligent data analysis and other methods of data analysis.
3. Shopping cart analysis task. Information table. Association rule.
4. Support and reliability of association rule. Types of association rules.
5. Search for binary association rules – general statement of the problem.
6. Support and reliability of the measure of association rules. Strong association rules. Other measures of association rule evaluation.
7. Frequent sets. Naive algorithm for finding frequent sets.
8. Basic algorithms for searching association rules. Apriori algorithm.
9. Monotonicity property of the support measure.
10. Dimensionality of the data under study. One-dimensional and multidimensional association rules.
11. Degree of abstraction of the data under study. Single-level and multilevel association rules.
12. Generation of candidate sets of frequent sets of different dimensions.
13. Varieties of the Apriori algorithm.

14. The idea of the FP Growth algorithm.
15. Sequence pattern detection problem. Sequence expressions, sequence length and size, maximum sequence. Finding frequent episodes in a sequence.
16. The basic GSP algorithm consists of the following steps: sorting, finding frequent events, finding all frequent sequences and maximal sequences.
17. Prefix Span algorithm.
18. Data types in cluster analysis. Real binary data.
19. Dependency measures in cluster analysis. Cosine dependence measure.
20. Iterative and hierarchical methods of cluster analysis. Graphical representation of clustering results.
21. DBSCAN algorithm.
22. Kohonen self-organizing maps and their use in cluster analysis.
23. Text mining.
24. Web mining.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Methods and algorithms of machine learning	Математического моделирования и анализа данных	Изменений не требуется	Протокол № 5 от 19.11.2024
Neural networks in machine learning	Математического моделирования и анализа данных	Изменений не требуется	Протокол № 5 от 19.11.2024
Methods for statistical analysis of complex data	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 5 от 19.11.2024
Data mining methods	Теории вероятностей и математической статистики	Изменений не требуется	Протокол № 5 от 19.11.2024

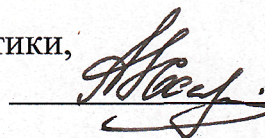
Заведующий кафедрой математического моделирования и анализа данных,
доктор экон. наук, доцент



В.И. Малюгин

19 ноября 2024 г.

Заведующий кафедрой теории вероятностей и математической статистики,
доктор физ.-мат. наук, профессор



А.Ю. Харин

19 ноября 2024 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ

Декан факультета
