

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе
и образовательным инновациям

О. Г. Прохоренко

«05» декабря 2023 г.

Регистрационный № УД – 1507/м.



ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

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

7-06-0533-05 Прикладная математика и информатика

Профилизация: Компьютерный анализ данных

2023 г.

Учебная программа составлена на основе ОСВО 7-06-0533-05-2023 специальности 7-06-0533-05 «Прикладная математика и информатика», примерного учебного плана, регистрационный № 7-06-05-016/пр от 18.01.2023 г., учебного плана №М53-5.3-43/уч. от 15.02.2023 г.

СОСТАВИТЕЛЬ:

Н.Н. Труш, профессор кафедры теории вероятностей и математической статистики Белорусского государственного университета, доктор физико-математических наук, профессор

РЕЦЕНЗЕНТ:

Н.В. Лапицкая, заведующая кафедрой программного обеспечения информационных технологий Белорусского государственного университета информатики и радиоэлектроники, кандидат технических наук, доцент

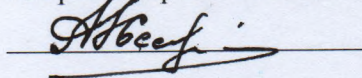
РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики БГУ
(протокол № 4 от 28.11.2023 г.)

Научно-методическим советом БГУ
(протокол № 3 от 30.11.2023 г.)

Заведующий кафедрой

теории вероятностей и математической статистики



А.Ю.Харин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

«Интеллектуальный анализ данных» – дисциплина, изучающая современные методы и алгоритмы визуализации, систематизации, исследования, анализа и прогнозирования данных большой размерности.

Основными целями дисциплины являются:

- 1) изучение современных методов и алгоритмов интеллектуального анализа данных;
- 2) формирование навыков решения практических задач с использованием современного программного обеспечения.

В рамках поставленных целей **задачи учебной дисциплины** состоят в следующем:

- 1) Изучение основных методов и алгоритмов Data Mining и их сравнительный анализ.
- 2) Развитие практических навыков использования основных методов и алгоритмов интеллектуального анализа данных.

Принципы изложения материала

В рамках первого раздела дисциплины на лекциях изучаются основные понятия анализа неструктурированных данных (Data Mining), на базе которых изучаются разделы, составляющие ее основное содержание. Рассматриваются основные проблемы, методы и алгоритмы интеллектуального анализа данных, включая поиск ассоциативных правил, поиск типовых последовательностей, методы кластерного анализа, анализа текстовых документов и страниц web. Рассматривается применение систем искусственного интеллекта и нейронных сетей для решения задач интеллектуального анализа данных. На лекциях излагается более сложный материал, вызывающий вопросы у магистрантов. В рамках выполнения лабораторных работ изучаются основные алгоритмы интеллектуального анализа данных, магистранты приобретают практические навыки решения конкретных задач с использованием пакета R или других программных средств.

Место учебной дисциплины в системе подготовки специалиста с углубленным высшим образованием.

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

Связи учебной дисциплины с другими дисциплинами

Основой для изучения дисциплины «Интеллектуальный анализ данных» является учебная дисциплина «Теория вероятностей и математическая

статистика». Дисциплина «Интеллектуальный анализ данных» будет полезна при прохождении практики и написании магистерских диссертаций.

Требования к компетенциям

Освоение учебной дисциплины «Интеллектуальный анализ данных» должно обеспечить формирование следующих компетенций:

универсальные компетенции:

УК-1. Применять методы научного познания в исследовательской деятельности, генерировать и реализовывать инновационные идеи.

Специализированные компетенции:

СК-2. Использовать знания вероятностных моделей и методов интеллектуального анализа данных.

В результате освоения учебной дисциплины магистрант должен:

знать:

- определение основных понятий интеллектуального анализа данных;
- определения различных мер расстояния;
- различные меры поиска документов и страниц web;

уметь:

- строить различные матрицы информации и пользоваться ими;
- вычислять различные меры расстояния;

владеть:

- основными методами интеллектуального анализа данных;
- знаниями различных мер расстояния в интеллектуальном анализе данных;
- основными алгоритмами интеллектуального анализа данных и их реализации.

Структура учебной дисциплины

Учебная дисциплина изучается во 2 семестре. Всего на изучение учебной дисциплины «Интеллектуальный анализ данных» отведено:

– для очной формы получения углубленного 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.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел I. Введение в интеллектуальный анализ данных

Тема 1.1. Data Mining. Что означает Data Mining? Методы и этапы интеллектуального анализа данных. Задачи и области применения Data Mining. Процесс открытия знаний. Отличие Data Mining от других методов анализа данных.

Раздел II. Методы поиска ассоциативных правил

Тема 2.1. Типы ассоциативных правил. Задача анализа корзины покупок. Таблица информации. Ассоциативное правило. Поддержка и достоверность ассоциативного правила. Типы ассоциативных правил.

Тема 2.2. Бинарные ассоциативные правила. Поиск бинарных ассоциативных правил – общая постановка задачи. Поддержка и достоверность меры ассоциативных правил. Сильные ассоциативные правила. Другие меры оценки ассоциативных правил. Наборы частые. Алгоритм наивный нахождения наборов частых. Основные алгоритмы поиска ассоциативных правил. Алгоритм Apriori. Свойство монотонности меры поддержки.

Тема 2.3. Многомерные и многоуровневые ассоциативные правила. Размерность исследуемых данных. Одномерные и многомерные ассоциативные правила. Степень абстракции исследуемых данных. Одноуровневые и многоуровневые ассоциативные правила. Генерация наборов кандидатов в наборы частые различной размерности. Разновидности алгоритма Apriori. Идея алгоритма FP Growth.

Раздел III. Поиск типовых последовательностей

Тема 3.1. Шаблоны последовательностей. Задача обнаружения шаблонов последовательностей. Выражения последовательности, длина и размер последовательности, максимальная последовательность. Нахождение частых эпизодов в последовательности.

Тема 3.2. Исследование типовых последовательностей. Основной алгоритм GSP состоящий из следующих шагов: сортировка, нахождение событий частых, нахождение всех частых последовательностей и максимальных последовательностей. Исследование типовых последовательностей. Алгоритм Prefix Span. Области применения.

Тема 3.3. Временные ряды. Анализ временных рядов. Основные определения и понятия. Примеры. Оценивание основных характеристик временных рядов. Параметрические модели временных рядов. Прогнозирование. Модели GARCH.

Раздел IV. Нейронные сети

Тема 4.1. Применение нейронных сетей. Нейронные сети. Модели нейронных сетей. Искусственный нейрон. Задачи Data Mining, решаемые с помощью нейронных сетей. Двухслойный персептрон. Примеры решения прикладных задач.

Раздел V. Кластерный анализ

Тема 5.1. Методы кластерного анализа. Типы данных в кластерном анализе. Вещественные бинарные данные. Меры зависимости. Косинусная мера зависимости. Алгоритмы и методы кластерного анализа данных.

Раздел VI. Анализ текстовых документов и страниц Web

Тема 6.1. Методы и алгоритмы анализа текстов и страниц Web. Поиск текстовых документов на основании запроса и на основе имеющего документа. Классификация текстовых документов. Группировка текстовых документов и анализ цитирования в сети. Меры оценки поиска документов. Матрица TFM. Методы и алгоритмы в анализе текстовых документов и страниц Web.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная форма получения углубленного высшего образования
с применением дистанционных образовательных технологий (ДОТ)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1		3	4	5	6	7	8	9
1	Введение в интеллектуальный анализ данных	2			2			
1.1	Data Mining	2			2			Устный опрос. Дискуссия
2	Методы поиска ассоциативных правил	6			6			
2.1	Типы ассоциативных правил	2			2			Контрольная работа
2.2	Бинарные ассоциативные правила	2			2			Отчет по лабораторной работе
2.3	Многомерные и многоуровневые ассоциативные правила	2			2			Коллоквиум
3	Поиск типовых последовательностей	6			6			
3.1	Шаблоны последовательностей	2			2			Контрольная работа
3.2	Исследование типовых последовательностей	2			2			Отчет по лабораторной работе
3.3	Временные ряды	2			2			Контрольная работа
4	Нейронные сети	2			2			

4.1	Применение нейронных сетей	2			2			Устный опрос. Дискуссия
5	Кластерный анализ	2			2			
5.1	Методы кластерного анализа	2			2			Отчет по лабораторной работе
6	Анализ тестовых документов и страниц Web	2			2			
6.1	Методы и алгоритмы анализа текстов и страниц Web	2			2			Контрольная работа
ИТОГО		20			20			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Труш, Н. Н. Введение в компьютерный и интеллектуальный анализ данных: учебные материалы для студентов специальностей 1-31 03 05 "Актарная математика", 1-98 01 01 "Компьютерная безопасность (математические методы и программные системы)", 1-31 03 06-01 "Экономическая кибернетика (математические методы и компьютерное моделирование в экономике)" / Н. Н. Труш; БГУ, Факультет прикладной математики и информатики, Кафедра теории вероятностей и математической статистики. - Минск: БГУ, 2022. - 69 с. - URL: <https://elib.bsu.by/handle/123456789/277034>.
2. Яцков, Н. Н. Интеллектуальный анализ данных [Электронный ресурс] : электронный учебно-методический комплекс для специальностей: 1-98 01 01 «Компьютерная безопасность» (по направлениям), 1-31 04 02 «Радиофизика», 1-31 04 03 «Физическая электроника» / Н. Н. Яцков ; БГУ, Фак. радиофизики и компьютерных технологий , Каф. системного анализа и компьютерного моделирования. - Минск : БГУ, 2023. - URL: <https://elib.bsu.by/handle/123456789/306842>.
3. Макшанов, А. В. Технологии интеллектуального анализа данных: учебное пособие / А. В. Макшанов, А. Е. Журавлев. - Санкт-Петербург; Москва; Краснодар: Лань, 2018. - 208 с. - URL: <https://e.lanbook.com/book/206711>.

Перечень дополнительной литературы

1. Труш Н.Н. Асимптотические методы статистического анализа временных рядов. – Мн. БГУ, 1999.
2. Olson D.L, Wu D. Prediktive Data Mining Models. – Springer. 2020. – 129 p.
3. Rutkowski L., Jaworski M., Dudo P. Stream Data Mining: Algorithms and Their Probabilistik Properties. – Springer. 2020. – 340 p.
4. Ghavami P. Big Data Analitics Methods. Analitics Techniques in Data Mining. Deep Learning and Natural Language Processing. – De Gruyter Pub., 2019. – 254 p.
5. Rutkowski L., Jaworski M., Dudo P. Stream Data Mining: Algorithms and Their Probabilistik Properties. Springer, 2020. 340 p.
6. Ширяев А.Н. Основы стохастической финансовой математики. Том 1, 2. – М.: Фазис, 1998.
7. Барсегян А.А. Анализ данных и процессов: учеб. пособие – Пер. с англ. под ред. В.Г.Трилиса. СПб.: БХВ-Петербург, 2009. – 512 с.
8. Чубукова И.А. Data Mining, М : НОУ “Интуит”, 2016. – 471 с.
9. 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.

10. Han J., Kamber M., Morgan Kaufman. Data Mining: concepts and techniques. – Morgan Kaufman Pub., 2012. – p.324.
11. Hand J., Mannila H., Smyth P. Principles of Data Mining. – MIT Press, 2001. – p.261.
12. Morzy T., Eksploatacja danych. Metody i algorytmy. – PWN, Warszawa, 2013. – 554 p.
13. Francq C., Zakoian J-M. GARCH models: structure, statistical inference and financial applications. – Wiley, 2010. – 489 p.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Объектом диагностики компетенций магистрантов являются знания, умения, полученные ими в результате изучения учебной дисциплины. Выявление учебных достижений магистрантов осуществляется с помощью мероприятий текущего контроля и промежуточной аттестации.

Для диагностики компетенций используется следующий диагностический инструментарий:

1. Устная форма: устный опрос, дискуссия, коллоквиум;
2. Устно-письменная форма: отчет по лабораторной работе, контрольная работа.

Формой промежуточной аттестации по учебной дисциплине «Интеллектуальный анализ данных» учебным планом предусмотрен **зачет**.

Примерная тематика лабораторных занятий

1. Реализация алгоритмов поиска ассоциативных правил.
Раздел 2. Методы поиска ассоциативных правил. Среда разработки по выбору студентов.

Форма контроля – письменный отчет по лабораторной работе.

2. Реализация алгоритмов поиска типовых последовательностей.
Раздел 3. Поиск типовых последовательностей. Среда разработки по выбору студентов.

Форма контроля – письменный отчет по лабораторной работе.

3. Реализация алгоритмов кластерного анализа. Тема 5.1. Методы кластерного анализа. Среда разработки по выбору студентов.

Форма контроля – письменный отчет по лабораторной работе.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используется **практико-ориентированный подход**, который предполагает

- освоение содержания учебного материала через решение практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;
- использование процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

При организации образовательного процесса используются следующие методы:

- **метод учебной дискуссии**, который предполагает участие студентов магистратуры в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме. Использование метода обеспечивает появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения.
- **метод группового обучения**, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы магистрантов по учебной дисциплине следует использовать современные информационные технологии, разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, презентации лекций, методические указания к лабораторным занятиям, материалы текущего контроля и промежуточной аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательного стандарта углубленного высшего образования и учебно-программной документации, в том числе вопросы для подготовки к зачету,

задания, вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов и др.).

Примерный перечень вопросов к зачету

1. Методы и этапы интеллектуального анализа данных.
2. Задачи и области применения Data Mining.
3. Таблица информации. Ассоциативное правило. Типы ассоциативных правил.
4. Поддержка и достоверность ассоциативного правила. Типы ассоциативных правил.
5. Поиск бинарных ассоциативных правил. Поддержка и достоверность меры ассоциативных правил.
6. Наборы частые. Основные алгоритмы поиска ассоциативных правил. Алгоритм Apriori.
7. Размерность исследуемых данных. Одномерные и многомерные ассоциативные правила.
8. Степень абстракции исследуемых данных. Одноуровневые и многоуровневые ассоциативные правила.
9. Задача обнаружения шаблонов последовательностей. Выражения последовательности, длина и размер последовательности, максимальная последовательность.
10. Нахождение частых эпизодов в последовательности.
11. Оценивание основных характеристик временных рядов.
12. Основной алгоритм GSP.
13. Исследование типовых последовательностей. Алгоритм Prefix Span. Области применения.
14. Анализ временных рядов. Основные определения и понятия. Модели GARCH.
15. Нейронные сети. Модели нейронных сетей. Искусственный нейрон.
16. Задачи Data Mining, решаемые с помощью нейронных сетей. Двухслойный персептрон.
17. Типы данных в кластерном анализе. Вещественные бинарные данные.
18. Меры зависимости. Косинусная мера зависимости. Алгоритмы и методы кластерного анализа данных.
19. Поиск текстовых документов на основании запроса и на основе имеющегося документа. Классификация текстовых документов.
20. Меры оценки поиска документов. Матрица TFM. Методы и алгоритмы в анализе текстовых документов и страниц Web.

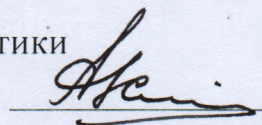
ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

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)
дисциплина не требует согласования			

Заведующий кафедрой

теории вероятностей и математической статистики

д.ф.-м.н., профессор



А.Ю.Харин

28.11.2023

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УО

на ____ / ____ учебный год

№№ Пп	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры теории вероятностей и математической статистики

(протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

(ученая степень, звание)

(подпись)

(И.О. Фамилия)

УТВЕРЖДАЮ

Декан факультета

(ученая степень, звание)

(подпись)

(И.О.Фамилия)