

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
Высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ
декан факультета _____ С.А.Баркалов
«31» августа 2021 года



РАБОЧАЯ ПРОГРАММА
дисциплины

«Интеллектуальный анализ данных»

Направление подготовки 38.03.05 БИЗНЕС-ИНФОРМАТИКА

Профиль Информационные системы в бизнесе

Квалификация выпускника бакалавр

Нормативный период обучения 4 года/4 года 11 м

Форма обучения очная/заочная

Год начала подготовки 2019

Автор программы _____ /Литвиненко Ю.В./

Заведующий кафедрой
систем
автоматизированного
проектирования и
информационных систем _____ /Львович Я.Е./

Руководитель ОПОП _____ /Наролина Т.С./

Воронеж 2021

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

формирование представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining) и методах их решения, которые помогают выявлять, формализовать и решать практические задачи анализа данных, возникающие в процессе профессиональной деятельности.

1.2. Задачи освоения дисциплины

- изучение основных понятий интеллектуального анализа данных;
- изучение основных задач Data Mining;
- изучение методов интеллектуального анализа данных;
- изучение основных этапов интеллектуального анализа;
- освоение инструментальных средств анализа данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Интеллектуальный анализ данных» относится к дисциплинам вариативной части блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Интеллектуальный анализ данных» направлен на формирование следующих компетенций:

ПК-13 - умение проектировать и внедрять компоненты ИТ-инфраструктуры предприятия, обеспечивающие достижение стратегических целей и поддержку бизнес-процессов

ДПК-1 - Способность создавать модели объектов и процессов экономических систем

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-13	знать основные задачи и методы интеллектуального анализа данных
	уметь формулировать задачи анализа данных
	владеть технологиями разработки алгоритмов и программными системами анализа данных
ДПК-1	знать принципы обработки больших массивов данных, способы их представления и хранения
	уметь выбирать адекватные алгоритмы их решения, оценивать качество получаемых решений
	владеть средствами автоматизации интеллектуального анализа и обработки данных

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Интеллектуальный анализ данных» составляет 5 з.е.

**Распределение трудоемкости дисциплины по видам занятий
очная форма обучения**

Виды учебной работы	Всего часов	Семестры
		7
Аудиторные занятия (всего)	72	72
В том числе:		
Лекции	36	36
Лабораторные работы (ЛР)	36	36
Самостоятельная работа	81	81
Часы на контроль	27	27
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость: академические часы	180	180
зач.ед.	5	5

заочная форма обучения

Виды учебной работы	Всего часов	Семестры
		7
Аудиторные занятия (всего)	16	16
В том числе:		
Лекции	6	6
Лабораторные работы (ЛР)	10	10
Самостоятельная работа	155	155
Контрольная работа		
Часы на контроль	9	9
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость: академические часы	180	180
зач.ед.	5	5

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

**5.1 Содержание разделов дисциплины и распределение
трудоемкости по видам занятий
очная форма обучения**

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Основы интеллектуального анализа данных	Понятие интеллектуального анализа данных. Data Mining как часть рынка информационных технологий. Набор данных и их атрибутов. Задачи Data Mining. Основы анализа данных.	6	8	25	39
2	Методы интеллектуального анализа данных	Задачи классификации и прогнозирования. Деревья решений. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация. Задача кластеризации. Алгоритм k-средних (k-means). Поиск ассоциативных правил. Задача визуализации.	20	24	31	75
3	Процесс применения интеллектуальных технологий	Основные этапы интеллектуального анализа. Инструментальные средства анализа данных. Применение Data Mining для решения бизнес-задач. Web Mining. Text Mining.	10	4	25	39

Итого	36	36	81	153
--------------	-----------	-----------	-----------	------------

заочная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Основы интеллектуального анализа данных	Понятие интеллектуального анализа данных. Задачи Data Mining.	2	4	50	56
2	Методы интеллектуального анализа данных	Задачи классификации и прогнозирования. Деревья решений. Метод опорных векторов. Метод "ближайшего соседа". Байесовская классификация. Задача кластеризации. Алгоритм k-средних (k-means). Поиск ассоциативных правил. Задача визуализации.	2	4	55	61
4	Процесс применения интеллектуальных технологий	Инструментальные средства анализа данных.	2	2	50	54
Итого			6	10	155	171

5.2. Перечень лабораторных работ очная форма обучения

1. Первичный анализ наборов данных.
2. Исследование метода главных компонент с помощью статистического пакета Statgraphics plus 5.0.
3. Обнаружение логических закономерностей на основе деревьев решений с помощью пакета Deductor Studio Lite.
4. Линейная регрессия
5. Наивный байесовский классификатор.
6. Проведение кластерного анализа с помощью статистического пакета Statgraphics plus 5.0.
7. Машина опорных векторов.
8. Ассоциативные правила.

заочная форма обучения

1. Первичный анализ наборов данных.
2. Исследование метода главных компонент с помощью статистического пакета Statgraphics plus 5.0.
3. Обнаружение логических закономерностей на основе деревьев решений с помощью пакета Deductor Studio Lite.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-13	знать основные задачи и методы интеллектуального анализа данных	Тест	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь формулировать задачи анализа данных	Решение стандартных практических задач	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть технологиями разработки алгоритмов и программными системами анализа данных	Решение прикладных задач в конкретной предметной области	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ДПК-1	знать принципы обработки больших массивов данных, способы их представления и хранения	Тест	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь выбирать адекватные алгоритмы их решения, оценивать качество получаемых решений	Решение стандартных практических задач	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	владеть средствами автоматизации интеллектуального анализа и обработки данных	Решение прикладных задач в конкретной предметной области	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 7 семестре для очной формы обучения, 7 семестре для заочной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
-------------	---	---------------------	---------	--------	--------	----------

ПК-13	знать основные задачи и методы интеллектуального анализа данных	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	уметь формулировать задачи анализа данных	Решение стандартных практических задач	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть технологиями разработки алгоритмов и программными системами анализа данных	Решение прикладных задач в конкретной предметной области	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
ДПК-1	знать принципы обработки больших массивов данных, способы их представления и хранения	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	уметь выбирать адекватные алгоритмы их решения, оценивать качество получаемых решений	Решение стандартных практических задач	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	владеть средствами автоматизации интеллектуального анализа и обработки данных	Решение прикладных задач в конкретной предметной области	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типичные контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1 Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

2 Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;

в) определение по известным характеристиками объекта значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

3 Задача кластеризации заключается в ...

а) нахождения частых зависимостей между объектами или событиями;

б) определения класса объекта по его характеристиками;

в) определение по известным характеристиками объекта значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

4 Целью поиска ассоциативных правил является ...

а) **нахождения частых зависимостей между объектами или событиями;**

б) определения класса объекта по его характеристиками;

в) определение по известным характеристиками объекта значение некоторого его параметра;

г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

5 Регрессия — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

6 Кластеризация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных.

7 Ассоциация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

8 Машинное обучение — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения

закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.

г) **подразделение искусственного интеллекта изучающий методы построения**

алгоритмов, способных обучаться на данных

9. Что из перечисленного не является алгоритмом добычи данных

а) классификация;

б) анализ последовательностей;

в) регрессия;

г) **корреляция;**

10. Что из перечисленного не является типом шкал измерений:

а) **абсолютная**

б) номинальная

в) порядковая

г) интервальная

7.2.2 Примерный перечень заданий для решения стандартных задач
(минимум 10 вопросов для тестирования с вариантами ответов)

7.2.3 Примерный перечень заданий для решения прикладных задач
(минимум 10 вопросов для тестирования с вариантами ответов)

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. Понятие интеллектуального анализа данных

2. Понятие знания

3. Данные и их атрибуты

4. Типы шкал измерений

5. Метаданные

6. Задачи Data Mining

7. Постановка задачи корреляционного анализа

8. Особенность регрессионного анализа

9. Задачи классификации и прогнозирования

10. Точность прогноза

11. Деревья решений

12. Метод опорных векторов

13. Метод "ближайшего соседа"

14. Байесовская классификация

15. Задача кластеризации

16. Сравнение классификации и кластеризации

17. Методы кластерного анализа

18. Методы объединения или связи

19. Метод Варда
20. Метод невзвешенного попарного среднего
21. Метод взвешенного попарного среднего
22. Определение количества кластеров
23. Алгоритм k-средних (k-means)
24. Применение кластерного анализа
25. Модификации алгоритмов кластерного анализа
26. Поиск ассоциативных правил
27. Характеристики ассоциативных правил
28. Задача визуализации
29. Методы визуализации
30. Основные этапы интеллектуального анализа
31. Извлечение знаний
32. Инструментальные средства анализа данных
33. Применение Data Mining для решения бизнес-задач

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Экзамен проводится по тест-билетам, каждый из которых содержит 10 вопросов и задачу. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом, задача оценивается в 10 баллов (5 баллов верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 6 баллов.

2. Оценка «Удовлетворительно» ставится в случае, если студент набрал от 6 до 10 баллов

3. Оценка «Хорошо» ставится в случае, если студент набрал от 11 до 15 баллов.

4. Оценка «Отлично» ставится, если студент набрал от 16 до 20 баллов.)

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Основы интеллектуального анализа данных	ПК-13, ДПК-1	Тест, защита лабораторных работ
2	Методы интеллектуального анализа данных	ПК-13, ДПК-1	Тест, защита лабораторных работ
3	Процесс применения интеллектуальных технологий	ПК-13, ДПК-1	Тест, защита лабораторных работ

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи

компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Чернышова Г.Ю. Интеллектуальный анализ данных / Саратовский государственный социально-экономический университет. – Саратов, 2012. – 92 с.

2. Мусаев А.А. Интеллектуальный анализ данных: учебное пособие. / А.А.Мусаев – СПб.: СПбГТИ(ТУ), 2018. – 56 с.

3. Айвазян С. А., Бухштабер В. М., Юньюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности. - М.: Финансы и статистика, 1989.

4. В. Дюк, А. Самойленко Data Mining. Учебный курс. - СПб.: Изд-во : Питер, 2001.

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Лицензионное ПО

LibreOffice

Ресурс информационно-телекоммуникационной сети «Интернет»

<http://www.edu.ru/>

Образовательный портал ВГТУ

Информационная справочная система

<http://window.edu.ru>

<https://wiki.cchgeu.ru/>

Современные профессиональные базы данных

ХабрХабр

Адрес ресурса: <https://habr.com/ru/>

Microsoft Developer Network

Адрес ресурса: <https://msdn.microsoft.com/ru-ru/>

Deductor Studio
Statgraphics plus

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Лекционная аудитория и аудитории для практических занятий, оснащённые мультимедийным демонстрационным оборудованием (проектор, экран, звуковоспроизводящее оборудование), обеспечивающим демонстрацию мультимедиа материалов.

Аудитории для лабораторных занятий, оснащенные компьютерами с лицензионным программным обеспечением с возможностью подключения к сети «Интернет» и доступом в электронную информационно образовательную среду университета.

Аудитории для самостоятельной работы, оборудованные техническими средствами обучения: персональными компьютерами с лицензионным программным обеспечением с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду университета

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Интеллектуальный анализ данных» читаются лекции, проводятся лабораторные работы.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать

	лекцию по соответствующей теме, ознакомится с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед экзаменом, экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.

6 Лист регистрации изменений

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
1	Внесены изменения в рабочие программы дисциплин в части состава используемого лицензионного программного обеспечения, современных профессиональных баз данных и справочных информационных систем	31.08.2020	
2	Внесены изменения в рабочие программы дисциплин в части состава используемого лицензионного программного обеспечения, современных профессиональных баз данных и справочных информационных систем	31.08.2021	
3	Актуализирован перечень литературы, необходимой для освоения дисциплины	31.08.2021	