

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Воронежский государственный технический университет»

УТВЕРЖДАЮ

Декан факультета информационных
технологий и компьютерной безопасности

/П.Ю. Гусев/

24.01.2023 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Интеллектуальный анализ данных»

Направление подготовки 09.03.02 Информационные системы и технологии

Профиль Информационные системы и технологии цифровизации

Квалификация выпускника бакалавр

Нормативный период обучения 4 года

Форма обучения очная

Год начала подготовки 2023

Автор программы

С.Ю. Белецкая

Заведующий кафедрой
Систем автоматизированного
проектирования и
информационных систем

Я.Е. Львович

Руководитель ОПОП

О.Г. Яскевич

Воронеж 2023

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Целью дисциплины является изучение основных классов задач анализа данных; формирование систематизированного представления о концепциях, моделях и методах интеллектуального анализа данных; получение практических навыков использования технологий интеллектуального анализа данных в информационных системах.

1.2. Задачи освоения дисциплины

- ознакомление студентов с основными классами задач анализа данных в информационных системах;
- изучение методов и алгоритмов интеллектуального анализа данных, а также их прикладных аспектов, связанных с моделированием и оптимизацией информационных процессов и систем;
- освоение основных приёмов построения моделей анализа данных;
- приобретение навыков использования современных инструментальных средств анализа данных;
- овладение навыками разработки математического и программного обеспечения для решения аналитических задач обработки информации.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Интеллектуальный анализ данных» относится к дисциплинам части, формируемой участниками образовательных отношений блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

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

ПК-1 - Способен анализировать входную информацию в процессе проектирования и разработки информационных систем

ПК-5 - Способен использовать модели и методы в процессе формализации и алгоритмизации поставленных задач

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-1	Знать основные классы задач анализа данных в информационных системах
	Уметь проводить аналитическую обработку информации в процессе проектирования и разработки информационных систем

	Владеть навыками решения задач анализа данных в информационных системах с использованием современных инструментальных средств
ПК-5	Знать модели и методы интеллектуального анализа данных, использующиеся при проектировании и эксплуатации информационных систем
	Уметь использовать технологию интеллектуального анализа данных в процессе формализации и алгоритмизации поставленных задач
	Владеть навыками разработки программного обеспечения для решения аналитических задач обработки информации

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Интеллектуальный анализ данных» составляет 7 з.е.

Распределение трудоемкости дисциплины по видам занятий

очная форма обучения

Виды учебной работы	Всего часов	Семестры	
		7	8
Аудиторные занятия (всего)	102	72	30
В том числе:			
Лекции	28	18	10
Лабораторные работы (ЛР)	74	54	20
Самостоятельная работа	114	36	78
Курсовой проект	+	+	
Часы на контроль	36	-	36
Виды промежуточной аттестации: зачет экзамен	+	+	+
Общая трудоемкость: академические часы зач.ед.	252 7	108 3	144 4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лек ц	Лаб зан.	СР С	Всего, час
1	Введение в интеллектуальный анализ данных. Понятие Data Mining	Основные задачи анализа данных в информационных системах. Классификация современных методов анализа данных. Особенности технологии Data Mining.	2	8	15	25
2	Корреляционный и регрессионный анализ.	Задачи корреляционного анализа. Этапы корреляционного анализа. Методики проверки значимости коэффициента корреляции. Основные задачи регрессионного анализа. Линейные и нелинейные регрессионные модели. Этапы регрессионного анализа. Парный и множественный регрессионный анализ. Использование метода наименьших квадратов для оценки параметров уравнений парной и множественной регрессии. Методы анализа качества регрессионной модели. Прикладные аспекты регрессионного анализа.	4	12	18	34
3	Кластерный анализ	Постановка задачи кластерного анализа. Этапы кластерного анализа. Метрики кластеризации, способы их определения. Классификация методов кластерного анализа. Итеративные методы. Иерархические методы. Оценка качества разбиения на классы.	4	8	15	27
4	Методы классификации данных	Задачи классификации данных и подходы к их решению. Классификация с использованием деревьев	4	12	16	32

		решений. Байесовская классификация. Метод опорных векторов. Классификация данных методом k ближайших соседей. Прикладные задачи классификации данных в информационных системах.				
5	Решение задач прогнозирования	Основные подходы к решению задач прогнозирования. Классификация задач прогнозирования. Основные этапы прогнозирования. Регрессионные модели прогнозирования. Понятие временного ряда. Основные показатели ряда динамики. Компоненты временного ряда. Методы исследования и прогнозирования временных рядов. Адаптивные методы прогнозирования	4	10	15	29
6	Нейросетевые технологии анализа данных	Понятие искусственной нейронной сети. Задачи, решаемые с использованием искусственных нейронных сетей. Структура и принципы функционирования искусственного нейрона. Модель Маккаллока-Питса. Типы искусственных нейронов. Классификация искусственных нейронных сетей. Однослойные и многослойные сети. Подходы к обучению искусственных нейронных сетей. Обучение с учителем. Обучение без учителя. Многослойный персептрон и его применение. Методы обучения многослойного персептрона. Алгоритм обратного распространения ошибки. Нейронные сети с самоорганизацией, направления их использования. Однослойные сети на базе нейронов ВТА. Сети и карты Кохонена. Методы обучения самоорганизующихся нейронных сетей. Нейронные сети встречного распространения: архи-	8	16	20	44

		тектура, обучение, принципы функционирования. Использование сетей встречного распространения для распознавания и сжатия данных. Принципы построения и функционирования рекуррентных нейронных сетей. Рекуррентные сети как ассоциативные запоминающие устройства. Автоассоциативная сеть Хопфилда. Сеть Хемминга. Принципы построения и функционирования свёрточных нейронных сетей. Этапы нейросетевого моделирования. Практическое использование нейросетевых технологий для решения задач обработки и анализа данных. Технология решения задач аппроксимации, классификации, распознавания, прогнозирования на основе искусственных нейронных сетей. Оценка качества нейросетевых моделей.				
7	Технологии и инструментальные средства анализа данных	Основные этапы анализа данных в информационных системах. Современные инструментальные средства анализа данных. Классификация и структура современных программных систем анализа данных. Состав и функциональные возможности систем. Особенности и сравнительный анализ систем	2	8	15	25
Итого			28	74	114	216

5.2 Перечень лабораторных работ

1. Решение задач регрессионного анализа в EXCEL. Парная и множественная регрессия.
2. Решение задач дисперсионного анализа в EXCEL Однофакторный и многофакторный дисперсионный анализ.
3. Анализ признаков и оценка их информативности в аналитической системе DEDUCTOR
4. Решение задач анализа данных в аналитической системе DEDUCTOR.
5. Изучение возможностей библиотек Python для анализа данных.

- Предварительная обработка данных.
6. Решение задач регрессионного анализа в Python с использованием библиотеки Scikit-learn.
 7. Решение задач кластеризации данных в Python с использованием библиотек Scikit-learn и SciPy.
 8. Решение задач прогнозирования в Python на основе библиотек Scikit-learn и Statsmodels.
 9. Нейросетевой анализ данных в Python на основе библиотек Keras и TensorFlow.
 10. Решение задач нейросетевого прогнозирования в Python.
 11. Решение задач нейросетевой классификации в Python.
 12. Распознавание графических объектов на основе свёрточных нейронных сетей.

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

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсового проекта в 7 семестре для очной формы обучения.

Обобщённая тематика курсового проекта: «Разработка математического и программного обеспечения анализа данных для различных прикладных областей»

Задачи, решаемые при выполнении курсового проекта:

- Разработка моделей анализа данных
- Построение алгоритмов анализа данных
- Разработка программного обеспечения для решения прикладных задач анализа данных

Курсовой проект включает в себя графическую часть и расчетно-пояснительную записку.

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-1	Знать основные классы задач анализа данных в информационных системах	Знание основных классов задач анализа данных в информационных системах. Ответы на теоретические вопросы при защите лабораторных работ и курсового проекта	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Уметь проводить аналитическую обработку информации в процессе проектирования и разработки информационных систем	Умение решать задачи классификации, распознавания и прогнозирования с использованием интеллектуальных методов анализа данных. Выполнение лабораторных работ и курсового проекта	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть навыками решения задач анализа данных в информационных системах с использованием современных инструментальных средств	Решение прикладных задач анализа данных в конкретной предметной области, выполнение лабораторных работ и курсового проекта.	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ПК-5	Знать модели и методы интеллектуального анализа данных, используемые при проектировании и эксплуатации информационных систем	Знание . моделей и методов Data Mining. Ответы на теоретические вопросы при защите лабораторных работ и курсового проекта	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Уметь использовать технологию интеллектуального анализа данных в процессе формализации и алгоритмизации поставленных задач	Знание технологии интеллектуального анализа данных. Выполнение лабораторных работ и курсового проекта	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть навыками разработки программного обеспечения для решения аналитических задач	Умение разрабатывать программное обеспечение	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

	обработки информации	печение для решения задач анализа данных. Выполнение лабораторных работ и курсового проекта	ренный в рабочих программах	смотренный в рабочих программах
--	----------------------	---	-----------------------------	---------------------------------

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 7, 8 семестре для очной формы обучения по двух/четырёхбалльной системе:

«зачтено»

«не зачтено» (для зачёта в 7 семестре)

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Зачтено	Не зачтено
ПК-1	Знать основные классы задач анализа данных в информационных системах	Тест	Выполнение теста на 70-100%	Выполнение менее 70%
	Уметь проводить аналитическую обработку информации в процессе проектирования и разработки информационных систем	Решение стандартных практических задач	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть навыками решения задач анализа данных в информационных системах с использованием современных инструментальных средств	Решение прикладных задач в конкретной предметной области	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
ПК-5	Знать модели и методы интеллектуального анализа данных, используемые при проектировании и эксплуатации информационных систем	Тест	Выполнение теста на 70-100%	Выполнение менее 70%
	Уметь использовать технологию интеллектуального анализа данных в процессе формализации и алгоритмизации поставленных задач	Решение стандартных практических задач	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть навыками разработки программного обеспечения для решения аналитических задач обработки информации	Решение прикладных задач в конкретной предметной области	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

или

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно» (для экзамена в 8 семестре)

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ПК-1	Знать основные классы задач анализа данных в информационных системах	Тест	Выполнение теста на 90- 100%	Выполнение теста на 80- 90%	Выполнение теста на 70- 80%	В тесте менее 70% правильных ответов
	Уметь проводить аналитическую обработку информации в процессе проектирования и разработки информационных систем	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 навыками решения задач анализа данных в информационных системах с использованием современных инструментальных средств	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
ПК-5	Знать модели и методы интеллектуального анализа данных, используемые при проектировании и эксплуатации информационных систем	Тест	Выполнение теста на 90- 100%	Выполнение теста на 80- 90%	Выполнение теста на 70- 80%	В тесте менее 70% правильных ответов
	Уметь использовать технологию интеллектуального анализа данных в процессе формализации и алгоритмизации поставленных задач	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 за-	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

				дачах		
	Владеть навыками разработки программного обеспечения для решения аналитических задач обработки информации	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 (типовые 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. Какое значение принимает накопленная частота последнего элемента выборки?

частота последнего элемента;

1;

100;

общее число элементов.

2. Какие значения может принимать коэффициент корреляции Пирсона:

от 0 до 1;

от -10 до 10;

от -1 до 1;

от 0 до 10.

3. Что является результатом регрессионного анализа?

расстояние регрессии;

уравнение регрессии;

матрица регрессии;

коэффициент регрессии.

4. Какая технология используется для оперативного анализа данных?

Data Mining;

OLTP;

DSS;

OLAP.

5. Какой метод применяется для обнаружения скрытых знаний в

больших объемах данных?

Data Mining;

дискриминантный анализ;

OLAP;

дисперсионный анализ.

6. Какие ряды применяются для группировки выборок большого объема из непрерывных генеральных совокупностей?

вариационный ряд;

статистический ряд;

интервальный вариационный ряд;

интервальный статистический ряд.

7. Коэффициент корреляции между двумя исследуемыми величинами равен -0.9. Что можно сказать о зависимости между данными величинами?

сильная прямая зависимость;

сильная обратная зависимость;

слабая обратная зависимость;

зависимость отсутствует.

8. Какой метод используется для разделения элементов выборки на множество групп в зависимости от выявленных внутренних связей между ними?

факторный анализ;

дисперсионный анализ;

дискриминантный анализ;

кластерный анализ.

9. В чем заключаются отличия задач классификации и кластеризации данных?

отличий нет, эти понятия являются синонимами;

при кластеризации заранее известны группы, к которым должен быть отнесен объект;

при классификации заранее известны группы, к которым должен быть отнесен объект;

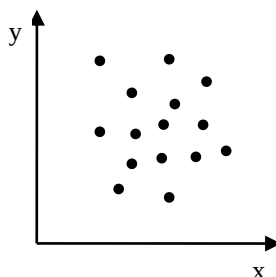
при классификации заранее не известны группы, к которым должен быть отнесен объект.

10. Какой метод применяется для исследования влияния на результативную переменную одного или нескольких качественных признаков?

факторный анализ;
дисперсионный анализ;
дискриминантный анализ;
кластерный анализ.

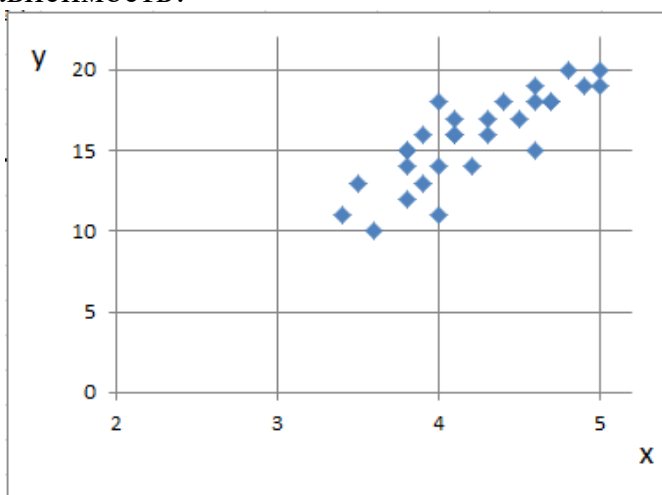
7.2.2 Примерный перечень заданий для решения стандартных задач

1. На рисунке изображено корреляционное поле, отражающее зависимость между исследуемыми величинами x и y . Что можно сказать о характере зависимости между данными величинами?



сильная прямая зависимость;
сильная обратная зависимость;
слабая обратная зависимость;
зависимость отсутствует.

2. На рисунке приведена зависимость $y(x)$. Какое уравнение регрессии описывает данную зависимость?



$y = -3.9 - 4.7x$;
 $y = -3.9 + 4.7x^2$;
 $y = -3.9 + 4.7x$;
 $y = -3.9 + 4.7 / x$.

3 Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

- А) должны быть качественно обоснованными
- Б) должны быть количественно измеримы
- В) не должны быть интеркоррелированы
- Г) должны быть функционально зависимыми

2 А, Б

3 Б, Г

4 В, Г

4 Объективными предпосылками появления ошибки модели регрессии явля

- А) нерепрезентативность выборки
- Б) вероятность ошибочной интерпретации переменных
- В) вероятность ошибочного измерения переменных
- Г) сложность модели

А, В

А, Б

Б, В

Б, Г

5 Значимость уравнения регрессии в целом оценивает

F-критерий Фишера

t- критерий Стьюдента

коэффициент детерминации r^2_{xy}

коэффициент корреляции r_{xy}

6 Коэффициент детерминации R^2

1 фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии факторов

2 фиксирует долю объясненной вариации результативного признака за счет не входящих в регрессии факторов

3 фиксирует долю объясненной вариации факторного признака за счет результативного признака

4 оценивает качество модели

7 Перечислите правильную последовательность этапов Knowledge Discovery in Databases –процесса обнаружения знаний в базах данных:

- 1 Трансформация, интерпретация результатов, выборка, очистка, построение моделей.
- 2 Построение моделей, выборка, очистка, трансформация, интерпретация результатов.
- 3 Построение моделей, выборка, очистка, трансформация, интерпретация результатов,
- 4 **Выборка, очистка, трансформация, построение моделей, интерпретация результатов.**

8 Какой из алгоритмов относится к алгоритмам кластеризации?

- 1 Алгоритм Дейкстры
- 2 **k-means – алгоритм**
- 3 Генетический алгоритм
- 4 Алгоритм Форда

9 Количество нейронов в слоях многослойного персептрона

- 1 Должны быть кратно двум
- 2 **Может быть произвольным**
- 3 Должно быть одинаковым
- 4 Может быть нулевым

10. В результате дисперсионного анализа был получен коэффициент корреляции, равный 0.9. Что можно сказать о характере зависимости результативного признака от факторного?

- 1 **Сильная прямая зависимость;**
- 2 Сильная обратная зависимость;
- 3 Слабая обратная зависимость;
- 4 Зависимость отсутствует.

7.2.3. Примерный перечень заданий для решения прикладных задач

1. Какой из методов Data Mining может использоваться для выявления

наборов часто покупаемых вместе продуктов?

- а) дерево решений;
- б) генетический алгоритм;
- в) поиск ассоциативных правил;**
- г) нечеткая логика.

2. Какой из методов Data Mining может использоваться для определения групп потребителей со схожими стереотипами поведения с целью предложения интересных именно им товаров?

- а) кластерный анализ;**
- б) генетический алгоритм;
- в) поиск ассоциативных правил;
- г) нечеткая логика.

3 Какой из методов Data Mining может использоваться для диагностики неисправностей на основе правил, описывающих сочетания различных симптомов?

- а) кластерный анализ;
- б) генетический алгоритм;
- в) поиск ассоциативных правил;
- г) дерево решений.**

4 При изучении зависимости $y=f(x_1, x_2, x_3)$ матрица парных коэффициентов корреляции оказалась следующей

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1	0,8	1	0,8	0,5
x_2	0,7	0,8	1	0,2
x_3	0,6	0,5	0,2	1

Какие факторы будут включены в уравнение множественной регрессии?

- 1 x_2, x_3**
- 2 x_1, x_3
- 3 x_1, x_2
- 4 x_1, x_2, x_3

Правильный ответ 1

5 В результате исследования взаимосвязи двух показателей получены следующие значения:

x_i	1	2	3
y_i	4	5	8

Тогда её интерполяционный многочлен второго порядка равен:

- 1) $x^2 - 3x + 6$
- 2) $x^2 - x + 4$
- 3) $x^2 - 2x + 5$
- 4) $x^2 - 4x + 7$

6 В результате исследования взаимосвязи двух показателей получены следующие пары значений:

x_i	1	2	3	4	5
y_i	5,3	6,3	4,8	3,8	3,3

Методом наименьших квадратов найти линейную функцию для приближения опытных (эмпирических) данных

- 1) $y = -0,65x + 6,65$
- 2) $y = -0,35x + 2,7$
- 3) $y = 0,25x + 4,5$
- 4) $y = 0,5x - 1,5$

7 Имеются данные по объёмам продаж торгового предприятия по месяцам

Ян-варь	Фев-раль	Март	Ап-рель	Май	Июнь	Июль	Август	Сен-тябрь	Ок-тябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

Используя метод наименьших квадратов (линейная функция), определить прогноз уровня продаж в ноябре.

- 1) 1,45
- 2) 1,28
- 3) 1,33
- 4) 1,54

8. К OLAP-системе сформулированы следующие требования:

- изменения в структуру измерений приходится вносить достаточно часто;
- есть иерархические измерения.

Какую модель данных лучше подходит для реализации системы:

- а) многомерная;
- б) реляционная, схема «звезда»;
- в) реляционная, схема «снежинка»;**
- г) логическая.

9. Какую роль может выполнять Excel в OLAP-системе?

- а) OLAP-сервер;
- б) OLAP-клиент;**
- в) СУБД;
- г) источник данных.

10 Имеются данные, характеризующие уровень инфляции в регионе:

Ян- варь	Фев- раль	Март	Ап- рель	Май	Июнь	Июль	Август	Сен- тябрь	Ок- тябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

Используя метод наименьших квадратов (линейная функция), определить прогноз уровня инфляции в декабре.

- 1)1,35
- 2)1,11**
- 3)1,43
- 4)1,24

7.2.4 Примерный перечень вопросов для подготовки к зачету

1. Основные задачи анализа данных в информационных системах.
2. Классификация современных методов анализа данных. Понятие о технологии Data Mining.
3. Технология корреляционного анализа.
4. Основные задачи регрессионного анализа. Линейные и нелинейные регрессионные модели. Этапы регрессионного анализа.
5. Парная и множественная регрессия. Оценка качества регрессионных моделей.
6. Постановка задачи кластерного анализа. Этапы кластерного анализа. Метрики кластеризации, способы их определения. Оценка качества разбиения на классы.
7. Классификация методов кластерного анализа. Итеративные методы.
8. Классификация методов кластерного анализа. Иерархические методы.
9. Задачи классификации данных и подходы к их решению. Прикладные задачи классификации данных в информационных системах
10. Классификация с использованием деревьев решений.
11. Байесовская классификация.
12. Метод опорных векторов.

13. Классификация данных методом k ближайших соседей.
14. Основные подходы к решению задач прогнозирования. Регрессионные модели прогнозирования.
15. Понятие временного ряда. Основные показатели ряда динамики. Компоненты временного ряда.
16. Методы исследования и прогнозирования временных рядов.
17. Адаптивные методы прогнозирования

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. Понятие искусственной нейронной сети. Задачи, решаемые с использованием искусственных нейронных сетей.
2. Структура и принципы функционирования искусственного нейрона. Модель Маккаллока-Питса.
3. Типы искусственных нейронов.
4. Классификация искусственных нейронных сетей. Однослойные и многослойные сети.
5. Подходы к обучению искусственных нейронных сетей. Обучение с учителем. Обучение без учителя.
6. Многослойный персептрон и его применение. Методы обучения многослойного персептрона. Алгоритм обратного распространения ошибки.
7. Нейронные сети с самоорганизацией, направления их использования. Однослойные сети на базе нейронов WTA. Сети и карты Кохонена. Методы обучения самоорганизующихся нейронных сетей.
8. Нейронные сети встречного распространения: архитектура, обучение, принципы функционирования. Использование сетей встречного распространения для распознавания и сжатия данных.
9. Принципы построения и функционирования рекуррентных нейронных сетей. Автоассоциативная сеть Хопфилда. Сеть Хемминга
10. Принципы построения и функционирования свёрточных нейронных сетей.
11. Технологии прогнозирования с использованием нейронных сетей.
12. Использование нейронных сетей для решения задач аппроксимации данных.
13. Решение задач нейросетевой классификации.
14. Этапы нейросетевого моделирования.
15. Технология нейросетевого моделирования с использованием библиотек Python.
16. Решение задач трансформации и визуализации данных.
17. Инструментальные средства анализа данных.

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Зачёт в 7 семестре 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 ре-

шение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. Зачёт ставится в случае, если студент набрал более 10 баллов.
2. Незачёт ставится в случае, если студент набрал менее 10 баллов

Экзамен в 8 семестре 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	Введение в интеллектуальный анализ данных. Понятие Data Mining	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
2	Корреляционный и регрессионный анализ.	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
3	Кластерный анализ	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
4	Методы классификации данных	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
5	Решение задач прогнозирования	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
6	Нейросетевые технологии анализа данных	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту
7	Технологии и инструментальные средства анализа данных	ПК-1, ПК-5	Тест, защита лабораторных работ, требования к курсовому проекту

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.

Защита курсовой работы, курсового проекта или отчета по всем видам практик осуществляется согласно требованиям, предъявляемым к работе, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 мин.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Питолин А.В. Нейросетевые технологии информационных систем : Учеб. пособие. - Воронеж : ВИВТ, 2007. - 129 с
2. Мусаев А.А. Интеллектуальный анализ данных: учебное пособие. / А.А.Мусаев – СПб.: СПбГТИ(ТУ), 2018. – 56 с.
3. Чеканов В.С., Кандаурова Н.В. Технологии обработки информации: учеб. пособие. – Ставрополь: Северо-Кавказский Федеральный университет, 2014. – 175 с. <http://www.iprbookshop.kr/63145/html>
4. Новикова Н.М. Обработка экспериментальных данных: Учеб. пособие. – Воронеж: ВГТУ, 2010. – 119 с.
5. Минаева Ю.В. Методы статистического и интеллектуального анализа данных: учеб. пособие. – Воронеж: ВГТУ, 2017. 90 с.

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:

Программное обеспечение

Microsoft Office Excel

Deductor Studio Academic

Python

Microsoft Visual Studio Community

Ресурсы информационно-телекоммуникационной сети «Интернет»

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

Образовательный портал ВГТУ

Информационные справочные системы

<http://window.edu.ru>

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

Современные профессиональные базы данных

ХабрХабр <https://habr.com/ru/>

Исходники.ru <https://sources.ru/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Специализированная лекционная аудитория, оснащенная оборудованием для лекционных демонстраций и проекционной аппаратурой
Дисплейный класс, оснащенный компьютерными программами для проведения лабораторного практикума

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

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

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.

Методика выполнения курсового проекта изложена в учебно-методическом пособии. Выполнять этапы курсового проекта должны своевременно и в установленные сроки.

Контроль усвоения материала дисциплины производится проверкой курсового проекта, защитой курсового проекта.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать 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, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоению 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 начинаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед зачетом, экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
----------	-----------------------------	----------------------------	--