

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

УТВЕРЖДАЮ

Декан факультета строительного

наименование факультета

/ Д.В. Панфилов /

И.О. Фамилия

31 августа 2021 г.

**РАБОЧАЯ ПРОГРАММА**

**дисциплины**

**«Модели и методы анализа больших данных»**

Направление подготовки 09.04.03 Прикладная информатика

Профиль Технологии искусственного интеллекта

Квалификация выпускника магистр

Нормативный период обучения 2 года

Форма обучения Очная

Год начала подготовки 2021 г.

Автор программы \_\_\_\_\_

  
подпись

А.А. Дробышев

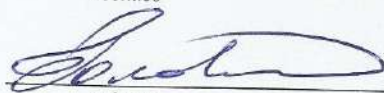
И.о. заведующего кафедрой  
инноватики и строительной физики  
имени профессора И.С. Суровцева

наименование кафедры, реализующей дисциплину

  
подпись

С.Н. Дьяконова

Руководитель ОПОП \_\_\_\_\_

  
подпись

П.А. Головинский

Воронеж 2021

# 1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

## 1.1. Цели дисциплины

Цель дисциплины «Модели и методы анализа больших данных» состоит в формировании знаний, умений и навыков (компетенций) по одному из приоритетных в современных информационных технологиях направлению - аналитической обработке больших данных. При освоении курса формируется представление о технических и методологических средствах анализа больших данных, обеспечивающих хранение и управление большими объемами данных, обеспечивается освоение эффективных методов организации неструктурированной информации, средств и способов работы с ней, генерации аналитических отчетов, внедрение прогностических моделей.

## 1.2. Задачи освоения дисциплины

Задачами освоения дисциплины «Модели и методы анализа больших данных» являются: ознакомление студентов с основными принципами машинного обучения; формирование представления о технических и методологических средствах анализа больших данных; формирование у студентов практических навыков сбора, обработки и анализа больших объемов данных, а также решения прикладных задач на языке Python.

# 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Модели и методы анализа больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений, блока Б1.

# 3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Модели и методы анализа больших данных» направлен на формирование следующих компетенций:

УК-4 - Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия;

ПК-4 - Способен управлять изменениями информационной среды, стратегией и рисками ИТ.

Компетенция	Результаты обучения, характеризующие сформированность компетенции
УК-4	знать: основы работ с хранилищами данных и NoSQL СУБД
	уметь: переводить проблемы предметной области на язык

	технологий обработки больших данных
	владеть: современными методами разработки технического, информационного и алгоритмического обеспечения Big Data.
ПК-4	знать: основные понятия, методы и средства в области Big Data; основы машинного обучения, визуализации и хранения больших данных
	уметь: применять подходы аналитической обработки больших данных при решении задач, связанных с управлением в сложных технических системах; осваивать методики использования программных средств, поддерживающих технологии Big Data для решения практических задач в предметной области
	владеть: методами обработки больших объемов данных, связанных с управлением в сложных технических системах с помощью современного ПО; способами применения современных теоретических и экспериментальных методов создания хранилищ данных

#### 4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Модели и методы анализа больших данных» составляет 6 з.е.

Распределение трудоемкости дисциплины по видам занятий

#### Очная форма обучения

Вид учебной работы	Всего часов	Семестры
		3
<b>Аудиторные занятия (всего)</b>	90	90
В том числе:		
Лекции	36	36
Практические занятия (ПЗ), в том числе в форме практической подготовки ( <i>при наличии</i> )	36	36
Лабораторные работы (ЛР), в том числе в форме практической подготовки ( <i>при наличии</i> )	18	18
<b>Самостоятельная работа</b>	90	90
Курсовой проект (работа) (есть, нет)	36	36

Контрольная работа (есть, нет)		
Вид промежуточной аттестации (зачет, зачет с оценкой, экзамен)		экзамен
Общая трудоемкость	час	216
	зач. ед.	6
		216
		6

## 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

### 5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

№ п/п	Наименование темы	Содержание раздела	Лекц	Прак зан.	Лаб. зан.	СРС	Всего, час
1	Сбор данных	Чтение файлов. Работа с текстовыми файлами. Извлечение данных из веб-ресурсов. Анализ кода HTML. Использование программных интерфейсов. Пример: Twitter API. Язык запросов SQL и NoSQL СУБД.	6	6	4	16	32
2	Работа с данными в пакете Pandas	Установка и использование пакета Pandas. Объекты библиотеки. Индексация и выбор данных. Операции с данными. Обработка отсутствующих данных. Снижение размерности. Иерархическая индексация. Объединение наборов данных. Агрегирование и группировка. Сводные таблицы. Векторизованные операции над строками. Работа с временными рядами.	6	6	4	16	32
3	Визуализация аналитических данных	Библиотека Matplotlib. Столбчатые диаграммы. Линейные графики. Диаграммы рассеяния. Визуализация погрешностей. График плотности и контурные графики. Настройки легенд на графиках. Построение трехмерных графиков. Разработка интерактивных информационных панелей. Отображение данных устройств, поступающих в реальном времени.	6	6	4	16	32
4	Стандартная аналитика	Поиск статистических зависимостей методами анализа статистических данных. Анализ и визуализация исторических данных.	6	6	2	14	28
5	Методы машинного обучения	Прогнозирование развития событий. Классификация текста. Распознавание и анализ изображений. Обнаружение аномалий и предсказание событий.	6	6	2	14	28
6	Агрегация и хранение данных	Организация ETL-процессов. Проектирование и разработка хранилищ данных. Поточковая обработка данных.	6	6	2	14	28
<b>Итого</b>			<b>36</b>	<b>36</b>	<b>18</b>	<b>90</b>	<b>180</b>

## **5.2 Перечень лабораторных работ**

1. Сбор и извлечение данных
2. Применение пакета Pandas для работы с данными
3. Основы работы с NoSQL СУБД и разработка баз данных
4. Визуализация анализа данных
5. Статистический анализ и прогнозирование
6. Методы кластеризации и классификации
7. Введение в машинное обучение
8. Введение в нейронные сети и глубокое обучение
9. Применение нейронных сетей в задачах анализа данных

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

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсового проекта в 3 семестре для очной формы обучения.

Примерная тематика курсового проекта: «Применение нейронной сети для решения задачи прогнозирования временного ряда»

Задачи, решаемые при выполнении курсового проекта:

- Выявление главных критериев и признаков, характеризующих данные;
- Выбор методики решения поставленной задачи;
- Разработка и применение информационно-математической модели для решения поставленной задачи.

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

Примерная тематика курсовых проектов:

1. Прогнозирование рейтинга художественного произведения при помощи методов машинного обучения
2. Применение нейронной сети для решения задачи прогнозирования временного ряда
3. Математические модели прогнозирования результатов соревнований
4. Использование нейронных сетей для поисковых систем интернет-магазинов
5. Использование нейросетей для анализа медицинских показателей
6. Прогнозирование товарооборота с помощью нейросетей
7. Применение методов машинного обучения для построения карты пути клиента
8. Применение методов машинного обучения в анализе содержания новостных порталов
9. Персональная характеристика и прогноз по данным социальных сетей с применением методов машинного обучения
10. Прогнозирование рейтинга районов города с применением методов машинного обучения

Учебным планом по дисциплине «Модели и методы анализа больших данных» не предусмотрено выполнение контрольной работы (контрольных работ).

## 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
УК-4	знать: основы работ с хранилищами данных и NoSQL СУБД	Посещение лекционных, лабораторных и практических занятий. Лабораторные работы, практические работы	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в срок, предусмотренный в рабочих программах. Практические задания решены	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок, предусмотренный в рабочих программах. Практические задания не решены
	уметь: переводить проблемы предметной области на язык технологий обработки больших данных	Посещение лекционных, лабораторных и практических занятий. Лабораторные работы, практические работы	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в срок, предусмотренный в рабочих программах. Практические задания решены	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок, предусмотренный в рабочих программах. Практические задания не решены
	владеть: современными методами разработки технического, информационного и алгоритмического обеспечения Big Data.	Посещение лекционных, лабораторных и практических занятий. Лабораторные работы, практические работы	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в срок, предусмотренный в рабочих программах. Практические задания решены	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок, предусмотренный в рабочих программах. Практические задания не решены
ПК-4	знать: основные понятия, методы и средства в области Big Data; основы машинного обучения, визуализации и хранения больших данных	Посещение лекционных, лабораторных и практических занятий. Лабораторные	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок,

		работы, практические работы	срок, предусмотренный в рабочих программах. Практические задания решены	предусмотренный в рабочих программах. Практические задания не решены
уметь: применять подходы аналитической обработки больших данных при решении задач, связанных с управлением в сложных технических системах; осваивать методики использования средств, поддерживающих технологии Big Data для решения практических задач в предметной области		Посещение лекционных, лабораторных и практических занятий. Лабораторные работы, практические работы	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в срок, предусмотренный в рабочих программах. Практические задания решены	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок, предусмотренный в рабочих программах. Практические задания не решены
владеть: методами обработки больших объемов данных, связанных с управлением в сложных технических системах с помощью современного ПО; способами применения современных теоретических и экспериментальных методов создания хранилищ данных		Посещение лекционных, лабораторных и практических занятий. Лабораторные работы, практические работы	Полное или частичное посещение лекционных, лабораторных и практических занятий. Выполнение работ в срок, предусмотренный в рабочих программах. Практические задания решены	Непосещение лекционных, лабораторных и практических занятий. Невыполнение работ в срок, предусмотренный в рабочих программах. Практические задания не решены

### 7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 3 семестре для очной формы обучения по четырехбалльной системе:

- «отлично»;
- «хорошо»;
- «удовлетворительно»;
- «неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
УК-4	знать: основы работ с хранилищами данных и NoSQL СУБД	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	уметь: переводить проблемы предметной области на язык технологий обработки больших 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: современными методами	Решение прикладных задач в	Задачи решены в полном объеме и	Продемонстрирован верный ход решения	Продемонстрирован верный ход решения в	Задачи не решены

	разработки технического, информационного и алгоритмического обеспечения Big Data.	конкретной предметной области	получены верные ответы	всех, но не получен верный ответ во всех задачах	большинстве задач	
ПК-4	знать: основные понятия, методы и средства в области Big Data; основы машинного обучения, визуализации и хранения больших данных	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	уметь: применять подходы аналитической обработки больших данных при решении задач, связанных с управлением в сложных технических системах; осваивать методики использования программных средств, поддерживающих технологии Big Data для решения практических задач в предметной области	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: методами обработки больших объемов данных, связанных с управлением в сложных технических системах с помощью современного ПО; способами применения современных теоретических и экспериментальных методов создания хранилищ 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 (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. Анализ данных включает в себя:



- а) Совокупность языков программирования, позволяющих строить сложные математические модели, на основе которых производится анализ данных;
- б) Совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности;
- в) Совокупность аксиом и основанных на них теорем, позволяющих извлекать из массива данных полезную для человечества информацию;
- г) Нет правильного ответа.

2. В чём заключается идея алгоритмов анализа данных, построенных на «прецедентах»?

- а) В протоколах «вход-выход», при числе элементов больше 1000 выходной сигнал не зависит от входного;
- б) Для всякого исследуемого массива данных можно подобрать математическую модель, хорошо их описывающую.;
- в) Похожие входные ситуации приводят к похожим выходным реакциям системы;
- г) Нет правильного ответа.

3. Алгоритмы, применяемые в задачах таксономии (отметьте лишний вариант).

- а) BIGFOR;
- б) FOREL;
- в) FOREL – 2;
- г) FOREVER;
- д) Лишних вариантов нет.

4. Суть алгоритма FOREL заключается в

- а) Определение центра масс множества точек многомерного пространства признаков, а также определение центра масс любой группы точек (таксона) из этого множества;
- б) Объединение в группы (таксоны) совокупности точек многомерного пространства признаков при выполнении критерия оптимальности объединения ( $F$ -критерия);
- в) Определении статистических характеристик (таких как макс., мин., сред. значения; дисперсия; среднее квадратическое отклонение и др.) множества точек, входящих в таксон;
- г) Нет правильного ответа.

5. Таксоны для алгоритма FOREL представляют собой:

- а) Многомерные эллипсоиды;
- б) Многомерные кубы;

- в) Многомерные сферы;
- г) Нет правильного ответа.

6. Первый пункт алгоритма FOREL после нормировки признаков объекта заключается в следующем:

- а) Вычисление радиуса гиперсферы  $R$ , охватывающей таксон множества точек;
- б) Определение числа таксонов, из которых состоит множество точек;
- в) Построение гиперсферы минимального радиуса  $R_0$ , которая охватывает всё множество точек;
- г) Нет правильного варианта

7. Какие из алгоритмов относятся к алгоритмам кластеризации

- а) Метод ближайшего соседа;
- б) Минимальное покрывающее дерево;
- в) k-Means алгоритм;
- г) Генетические алгоритмы.

8. Классификация методов Data Mining по задачам

- а) задачи классификации;
- б) задача кластеризации;
- в) задачи прогнозирования;
- г) все.

9. Этапы обучения в правильном порядке:

- 1) Сбор данных;
- 2) Подготовка данных (фильтрация, дополнение, кодирование);
- 3) Постановка задачи анализа;
- 4) Подбор параметров модели и алгоритма обучения;
- 5) Выбор модели (алгоритма анализа данных);
- 6) Обучение модели (автоматический поиск остальных параметров модели);
- 7) Анализ качества обучения, если неудовлетворительный переход на п. 5 или п. 4;
- 8) Анализ выявленных закономерностей, если неудовлетворительный переход на п. 1, 4 или 5.

- а) 3,1,2,5,4,6,7,8;
- б) 1,2,3,4,5,6,7,8;
- в) 2,3,1,5,6,8,7,4;
- г) нет правильного ответа.

10. Агрегация данных это

- а) процесс сбора данных;
- б) процесс обработки данных;
- в) процесс шифрования данных;

г) процесс записи данных.

## 7.2.2 Примерный перечень заданий для решения стандартных задач

1. Каков результат выполнения кода:

```
import pandas as pd
df= pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
print(df)
```

- а) строка [2, 1, 3];
- б) строка [6, 7, 8];
- в) матрица:

	one	two
0	2	6
1	1	7
2	3	8

г) матрица:

	one	two
0	6	2
1	7	1
2	8	3

2. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3, 5],
                  'two': [6, 7, 8, 9]})
print(df['two'].loc[1])
```

- а) 7;
- б) 8;
- в) 2;
- г) 1.

3. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3, 5],
                  'two': [6, 7, 8, 9]})
print(df.loc[1, 'two'])
```

- а) 2;
- б) 6;

в) 7;      г) 8.

4. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
df['one']=sorted(df['one'])
print(df)
```

а) матрица:

	one	two
0	1	6
1	2	7
2	3	8

б) матрица:

	one	two
0	6	2
1	7	1
2	8	3

в) 1

г) 0

5. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3, 5],
                  'two': [6, 7, 8, 9]})
S=df['two'].sum()
print(S)
```

а) 0;

б) 1;

в) 8;

г) 30.

6. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
print(df[['one']])
```

а) столбец:

[6  
7  
8]

б) столбец:

[2  
1  
3]

в) строка:

[2, 1, 3]

г) строка:

[6, 7, 8]

7. Каков результат выполнения кода:

```
import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
```

```

        'two': [6, 7, 8]])
print(df.iloc[0:2])

```

а) строка:  
[2, 6]

б) строка:  
[1, 7]

в) строки:  
[2, 6]  
[1, 7]

г) строки:  
[1, 7]  
[3, 8]

8. Каков результат выполнения кода:

```

import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
print(df.iloc[2])

```

а) строка:  
[2, 6]

б) строка:  
[3, 8]

в) строки:  
[2, 6]  
[1, 7]

г) строки:  
[1, 7]  
[3, 8]

9. Каков результат выполнения кода:

```

import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
df['two'].iloc[1]=20
print(df)

```

а) матрица:

	one	two
0	2	6
1	1	20
2	3	8

б) матрица:

	one	two
0	6	2
1	7	1
2	8	3

в) 20

г) 0

10. Каков результат выполнения кода:

```

import pandas as pd
df = pd.DataFrame({'one': [2, 1, 3],
                  'two': [6, 7, 8]})
df.loc[3]=[1, 1]
print(df)

```

а) матрица:

	one	two
0	2	6
1	1	20
2	3	8
3	1	1

б) матрица:

	one	two
0	6	2
1	7	1
2	8	3

в) строка:

[1,1]

г) 0

### 7.2.3 Примерный перечень заданий для решения прикладных задач

1. Приведены данные продаж на конец недели. Данные записывались в течение 6 недель. Составить прогноз на 7-ю и 8-ю неделю.

№ недели	1	2	3	4	5	6
объем	10	12	9	13	12	15

2. Провести классификацию  $n=6$  объектов, каждый из которых характеризуется двумя признаками.

Номер объекта	1	2	3	4	5	6
$x_{i1}$	5	6	5	10	11	10
$x_{i2}$	10	12	13	9	9	7

3. Дана табличная зависимость теплоемкости оксида углерода от температуры

$T, K$	300	400	500	600	700	800	900	1000
$C_p, \frac{Дж}{моль \cdot K}$	6.97	7.01	7.12	7.28	7.45	7.62	7.79	7.93

Построить аппроксимирующий полином первой степени.

4. Значения двух случайных величин приведены в таблице. Рассчитать коэффициент корреляции между этими величинами.

X	15	12	10	11	9	7	5
Y	1	2	4	3	6	4	2

5. При уровне значимости 0.05 методом дисперсионного анализа проверить нулевую гипотезу о влиянии фактора на качество объекта на основании пяти измерений для трех уровней факторов  $\Phi_1, \Phi_2, \Phi_3$

Номер измерения	Ф1	Ф2	Ф3
1	18	24	36
2	28	36	12
3	12	28	22
4	14	40	45
5	32	16	40

6. Имеются данные о динамике средних цен на вторичном рынке жилья по Невскому району Санкт-Петербурга, 2006г. (руб. за кв.м):

Месяц	Цена 1 кв.м (руб.)
январь	29 916,8
февраль	30 512,4
март	30 891,0
апрель	32 204,9
май	34 595,5
июнь	38 540,1
июль	46 204,5
август	55 273,9
сентябрь	60 806,8
октябрь	63 667,6
ноябрь	63 219,6
декабрь	62 338,7

На основе соответствующих данных определить основную тенденцию ряда динамики.

7. В таблице представлены данные по заработной плате сотрудников отдела. Рассчитать среднюю заработную плату и медианное значение.

№	1	2	3	4	5	6	7
Заработная плата, тыс. руб.	20	20	25	28	120	200	35

8. Дан закон распределения случайной величины  $X$ :

$x_i$	0	1	2	3
$p_i$	0.2	0.3	0.4	0.1

Вычислите математическое ожидание, дисперсию и среднее квадратическое отклонение.

9. Вероятностный прогноз для величины  $X$  -процентного изменения стоимости акций по отношению к их текущему курсу в течение 6 месяцев дан

в виде закона распределения:

$x_i$	5	10	15	20	25	30
$p_i$	0.1	0.1	0.3	0.2	0.2	0.1

Найти вероятность того, что покупка акций будет более выгодна, чем помещение денег на банковский депозит под 3% в месяц сроком на 6 месяцев.

10. Имеется набор из 8 точек в двумерном пространстве. Разделить этот набор данных на 2 кластера.

	X	Y
A	1	3
B	3	3
C	4	3
D	5	3
E	1	2
F	4	2
G	1	1
H	2	1

#### **7.2.4 Примерный перечень вопросов для подготовки к зачету** Не предусмотрено учебным планом.

#### **7.2.5 Примерный перечень вопросов для подготовки к экзамену**

1. Чтение файлов.
2. Работа с текстовыми файлами.
3. Извлечение данных из веб-ресурсов.
4. Анализ кода HTML.
5. Использование программных интерфейсов.
6. Пример: Twitter API.
7. Язык запросов SQL.
8. Установка и использование пакета Pandas.
9. Объекты библиотеки Pandas.
10. Индексация и выбор данных Pandas.
11. Операции с данными Pandas.
12. Обработка отсутствующих данных Pandas.
13. Снижение размерности Pandas.
14. Иерархическая индексация Pandas.
15. Объединение наборов данных Pandas.
16. Агрегирование и группировка Pandas.
17. Сводные таблицы Pandas.
18. Векторизованные операции над строками Pandas.
19. Работа с временными рядами Pandas.



20. Библиотека Matplotlib.
21. Столбчатые диаграммы Matplotlib.
22. Линейные графики Matplotlib.
23. Диаграммы рассеяния Matplotlib.
24. Визуализация погрешностей Matplotlib.
25. График плотности и контурные графики Matplotlib.
26. Настройки легенд на графиках Matplotlib.
27. Построение рекурсивных графиков Matplotlib.
28. Разработка интерактивных информационных панелей.
29. Отображение данных устройств, поступающих в реальном времени.
30. Поиск статистических зависимостей методами анализа статистических данных.
31. Анализ и визуализация исторических данных.
32. Поиск статистических зависимостей методами анализа статистических данных.
33. Анализ и визуализация исторических данных.
34. Прогнозирование развития событий.
35. Классификация текста.
36. Распознавание и анализ изображений.
37. Обнаружение аномалий и предсказание событий.
38. Организация ETL-процессов.
39. Проектирование и разработка хранилищ данных.
40. Поточковая обработка данных.

#### **7.2.6. Методика выставления оценки при проведении промежуточной аттестации**

Экзамен проводить в устно-письменной форме, которая включает ответы экзаменуемого на теоретические вопросы и решение им задач. 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	Сбор данных	УК-4, ПК-4	Тест, решение задач, защита

			лабораторных работ, защита курсового проекта
2	Работа с данными в пакете Pandas	УК-4, ПК-4	Тест, решение задач, защита лабораторных работ, защита курсового проекта
3	Визуализация аналитических данных	УК-4, ПК-4	Тест, решение задач, защита лабораторных работ, защита курсового проекта
4	Стандартная аналитика	УК-4, ПК-4	Тест, решение задач, защита лабораторных работ, защита курсового проекта
5	Методы машинного обучения	УК-4, ПК-4	Тест, решение задач, защита лабораторных работ, защита курсового проекта
6	Агрегация и хранение данных	УК-4, ПК-4	Тест, решение задач, защита лабораторных работ, защита курсового проекта

### **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 курсового проекта и 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. Пальмов, С.В. Интеллектуальный анализ данных [Электронный ресурс]: Учебное пособие / С. В. Пальмов. - Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. - 127 с. - ISBN 2227-8397. URL: <http://www.iprbookshop.ru/75376.html>
2. Головинский, П.А. Интеллектуальные информационные системы: теоретические основы и приложения [Текст] / Воронеж. гос. архитектур.-строит. ун-т. - Воронеж: Цифровая полиграфия, 2015 (Воронеж: ООО "Цифровая полиграфия", 2015). - 204 с.: ил. - ISBN 978-5-906384-26-3: 100-00.
3. Симчера, В. М. Методы многомерного анализа статистических данных : Учебное пособие / Симчера В. М. - Москва : Финансы и статистика, 2014. - 400 с. - ISBN 978-5-279-03184-9. URL: <http://www.iprbookshop.ru/18820.html>
4. Самостоятельная работа студентов [Текст] : методические указания к организации самостоятельной работы студентов направления 09.04.03 "Прикладная информатика" (магистерская программа "Технология искусственного интеллекта") всех форм обучения / ФГБОУ ВО "Воронеж. гос. техн. ун-т", кафедра инноватики и строительной физики ; сост. : П. А. Головинский, А. О. Шаталова. - Воронеж : Воронежский государственный технический университет, 2021. - 39 с. : табл.

## **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:**

1. Microsoft Office Power Point 2013/2007
2. Microsoft Office Excel 2013/2007
3. Adobe Acrobat Reader
4. Skype
5. Python
6. Яндекс.Браузер
7. <http://www.edu.ru/> – федеральный портал «Российское образование»
8. <https://education.cchgeu.ru> – образовательный портал ВГТУ
9. <http://window.edu.ru/> – единое окно доступа к информационным ресурсам
10. <https://cyberleninka.ru/> – научная электронная библиотека
11. <http://www.cmmarket.ru/> – мировые товарные рынки
12. <http://www.multistat.ru/> – многофункциональный статистический портал
13. <http://www.iprbookshop.ru> – электронно-библиотечная система
14. <http://biblioclub.ru> – ЭБС «Университетская библиотека онлайн»

## **9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА**

Для использования презентаций при проведении лекционных занятий необходимы аудитории, оснащенные презентационным оборудованием (компьютер с ОС Windows и программой PowerPoint или Adobe Reader, мультимедийный проектор и экран).

Для обеспечения практических и лабораторных занятий требуется компьютерный класс с комплексом лицензионного программного обеспечения: пакетами Microsoft Office, Python.

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

По дисциплине «Модели и методы анализа больших данных» читаются лекции, проводятся практические занятия и 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.

Практические занятия направлены на приобретение практических навыков расчета. Занятия проводятся путем решения конкретных задач в аудитории.

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. Этапы курсового проекта должны выполняться 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; пометить 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.
Практическое занятие	Конспектирование рекомендуемых источников. Работа с конспектом лекций, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы. Прослушивание аудио- и видеозаписей по заданной теме, выполнение расчетно-графических заданий, решение задач по алгоритму.
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: <ul style="list-style-type: none"> <li>- работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций;</li> <li>- выполнение домашних заданий и расчетов;</li> <li>- работа над темами для самостоятельного изучения;</li> <li>- участие в работе студенческих научных конференций, олимпиад;</li> <li>- подготовка к промежуточной аттестации.</li> </ul>
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 экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.