

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

УТВЕРЖДАЮ
Декан факультета  Д.В.Панфилов
«___» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА
дисциплины

«Модели и методы анализа больших данных»

Направление подготовки 09.04.03 Прикладная информатика

Профиль «Технологии искусственного интеллекта»

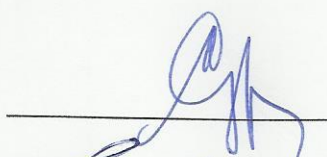
Квалификация выпускника магистр

Нормативный период обучения 2 года

Форма обучения очная

Год начала подготовки 2020

Автор программы  / Дробышев А.А.

Заведующий кафедрой
Инноватики и строительной
физики  / Суровцев И.С.

Руководитель ОПОП  / Головинский П.А.

Воронеж 20__

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 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 5 з.е.

Распределение трудоемкости дисциплины по видам занятий

Виды учебной работы	Всего часов	Семестры
		3
Аудиторные занятия (всего)	90	90
В том числе:		
Лекции	36	36
Практические занятия (ПЗ)	36	36
Лабораторные работы (ЛР)	18	18
Самостоятельная работа	54	54
Курсовой проект	+	+
Часы на контроль	36	36
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

№ 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	8	24
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	8	24
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	8	24
4	Стандартная аналитика	Поиск статистических зависимостей методами анализа статистических данных. Анализ и визуализация исторических данных.	6	6	2	10	24
5	Методы машинного обучения	Прогнозирование развития событий. Классификация текста. Распознавание и анализ изображений. Обнаружение аномалий и предсказание событий.	6	6	2	10	24
6	Агрегация и хранение данных	Организация ETL-процессов. Проектирование и разработка хранилищ данных. Поточковая обработка данных.	6	6	2	10	24
Итого			36	36	18	54	144

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 включают 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
УК-4	знать:	Лабораторные	Выполнение	Невыполнение

	основы работ с хранилищами данных и NoSQL СУБД	работы. Практические занятия	работ в срок, предусмотренный в рабочих программах	е работ в срок, предусмотренный в рабочих программах
	уметь: переводить проблемы предметной области на язык технологий обработки больших данных	Лабораторные работы. Практические занятия	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	владеть: современными методами разработки технического, информационного и алгоритмического обеспечения Big Data.	Лабораторные работы. Практические занятия	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ПК-4	знать: основные понятия, методы и средства в области Big Data; основы машинного обучения, визуализации и хранения больших данных	Лабораторные работы. Практические занятия	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	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	Лабораторные работы. Практические занятия	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	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	Лабораторные работы. Практические занятия	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 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 разработки технического, информационного и алгоритмического обеспечения Big Data.	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-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	Решение стандартных	Задачи решены в	Продемонстрирован	Продемонстрирован	Задачи не решены

	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	х практическ их задач	полном объеме и получены верные ответы	верный ход решения всех, но не получен верный ответ во всех задачах	верный ход решения в большинстве задач	
	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) Определении 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; среднеквадратическое 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:

б) матрица:

	one	two
0	1	6
1	2	7
2	3	8

	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 Φ_1 , Φ_2 , Φ_3

Номер измерения	Φ_1	Φ_2	Φ_3
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 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 рехмерных 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.

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Сбор данных	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ, защита реферата, требования к курсовому проекту
2	Работа с данными в пакете Pandas	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ, защита реферата, требования к курсовому проекту
3	Визуализация аналитических данных	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ, защита реферата, требования к курсовому проекту
4	Стандартная аналитика	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ, защита реферата, требования к курсовому проекту
5	Методы машинного обучения	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ, защита реферата, требования к курсовому проекту
6	Агрегация и хранение данных	УК-4, ПК-4	Тест, контрольная работа, защита лабораторных работ,

			защита реферата, 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

8.1.1 Основная литература

1. 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>
2. 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>
3. Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография / Л.А. Мыльников, Б. Краузе, М.

Кютц, К. Баде, И.А. Шмидт. - Москва: БИБЛИО-ГЛОБУС, 2017. - 334 с.: табл., граф., схем., ил. - Библиогр.: с. 311-328. - ISBN 978-5-9500501-7-6. URL: <http://biblioclub.ru/index.php?page=book&id=499006>

4. 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.

8.1.2 Дополнительная литература

1. Секлетова, Н.Н. Системный анализ и принятие решений [Электронный ресурс]: Учебное пособие / Н. Н. Секлетова, А. С. Тучкова. - Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. - 83 с. - ISBN 2227-8397. URL: <http://www.iprbookshop.ru/75407.html>

2. Интеллектуальные системы: учебное пособие / А. Семенов. - Оренбург: ОГУ, 2013. - 236 с. URL: <http://biblioclub.ru/index.php?page=book&id=259148>

3. Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография / Л.А. Мыльников, Б. Краузе, М. Кютц, К. Баде, И.А. Шмидт. - Москва: БИБЛИО-ГЛОБУС, 2017. - 334 с.: табл., граф., схем., ил. - Библиогр.: с. 311-328. - ISBN 978-5-9500501-7-6. URL: <http://biblioclub.ru/index.php?page=book&id=499006>

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. Использование презентаций при проведении лекционных занятий.
2. Консультирование посредством электронной почты.
3. <http://window.edu.ru/> - Единое окно доступа к информационным ресурсам.
4. <https://khashtamov.com/ru/pandas-introduction/> - Введение в pandas: анализ данных на Python.

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 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: - 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 три дня эффективнее всего использовать для повторения и систематизации материала.