

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)

Институт экономики и туризма

(Наименование института)

УТВЕРЖДАЮ:

Директор института



Козлов Д.А.

«11» сентября 2023 года

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Введение в Big Data

(НАИМЕНОВАНИЕ ДИСЦИПЛИНЫ)

направление подготовки / специальность

01.03.05. СТАТИСТИКА

(код и наименование направления подготовки (специальности))

направленность (профиль) подготовки

«БИЗНЕС-АНАЛИТИКА»

(направленность (профиль) подготовки))

г. Владимир

2023

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Введение в Big Data» является:

1. Изучение математических методов и подходов, используемых в программных системах обработки и анализа больших данных.

2. Развитие профессиональных навыков обучающихся за счет получения практического опыта работы с IT решениями в части обработки и анализа больших данных.

Задачи: усвоение знаний о сущности, структуре и видах математических моделей принятия решений; развитие навыков создания и решения моделей, необходимых в сфере управления.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Введение в Big Data» относится к базовой части учебного плана по направлению 01.03.05 Статистика.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОПОП (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	
ОПК-2. Способен формировать упорядоченные сводные массивы статистической информации и осуществлять расчет сводных и производных показателей в соответствии с утвержденными методиками, в том числе с применением необходимой вычислительной техники и стандартных компьютерных программ	ОПК-2.1. Знает методики формирования упорядоченных массивов статистической информации для решения профессиональных задач ОПК-2.2. Умеет применять современные информационные технологии и программные средства, для формирования массивов статистической информации ОПК-2.3. Владеет навыками расчета сводных и производных показателей для решения практических задач профессиональной деятельности	Знает методики работы с большими массивами информации для решения профессиональных задач Умеет применять современные информационные технологии и программные средства, для работы с большими массивами информации для решения профессиональных задач Владеет навыками расчета сводных и производных показателей при работе с большими массивами информации для решения профессиональных задач	Рейтинг-контроль, тесты, семинары, эссе.
ОПК-4. Способен понимать принципы работы	ОПК-4.1. Знает принципы работы современных информационных технологий	Знает принципы работы с большими данными Умеет выбирать	Рейтинг-контроль, тесты, семинары, эссе.

современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-4.2. Умеет выбирать информационные технологии ОПК-4.3. Владеет навыками использования современных информационных технологий при решении задач профессиональной деятельности	информационные технологии для работы с большими данными Владеет навыками использования технологий работы с большими данными	
---	--	--	--

4. ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 4 зачетные единицы, 144 часа.

Тематический план форма обучения – очная (срок обучения 4 года)

№ п/п	Наименование тем и/или разделов/тем дисциплины	Семестр	Неделя семестра	Контактная работа обучающихся с педагогическим работником				Самостоятельная Работа	Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные работы	в форме практической подготовки		
1	Методы многомерного статистического анализа и анализа нечисловой информации	5	1-7	4	4	4		15	Рейтинг-контроль №1
2	Технологии хранения и обработки больших данных	5	8-12	4	4	4		15	Рейтинг-контроль №2
3	Программирование обработки и загрузки больших данных	5	13-15	4	4	4		15	
4	Аналитика в больших данных	5	16-18	6	6	6		18	Рейтинг-контроль №3
Всего за 5 семестр:				18	18	18		63	Экзамен (27)
Наличие в дисциплине КП/КР									
Итого по дисциплине				18	18	18		63	Экзамен (27)

Содержание лекционных занятий по дисциплине

Тема 1. Методы многомерного статистического анализа и анализа нечисловой информации Факторный анализ. Дискриминантный анализ. Кластерный анализ. Многомерное шкалирование. Методы контроля качества.

Тема 2. Технологии хранения и обработки Больших данных Основные направления развития методов обработки и хранения данных. Volume. Закон Мура. Velocity. Variety.

Фреймворк Hadoop. Проблема хранения неструктурированных данных. Проблема преобразования данных. Семантические анализаторы. Самообучающиеся автоматы.

Тема 3. Программирование обработки и загрузки Больших данных 9 языков для Big Data (R, Python, Julia, Java, Scala, MATLAB, Go, Kafka, Hadoop). Фреймворки (Hadoop, Spark, Storm). Базы данных (Hive, Impala, Presto, Drill). Аналитические платформы (Rapid Miner, IBM SPSS Modeler, KNIME, Qlik Analytics Platform, STATISTICA Data Miner, Informatica Intelligent Data Platform, World Programming System, Deductor, SAS Enterprise Miner). Прочие инструменты (Zookeeper, Flume, IBM Watson Analytics, Dell EMC Analytic Insights Module, Windows Azure HDInsight, Microsoft Azure Machine Learning, Pentaho Data Integration, Teradata Aster Analytics, SAP BusinessObjects Predictive Analytics, Oracle Big Data Preparation).

Тема 4. Аналитика в больших данных Аналитика Big Data — реалии и перспективы в России и мире. Технологии и методы анализа, которые используются для анализа Big Data (Data Mining; краудсорсинг; смешение и интеграция данных; машинное обучение; искусственные нейронные сети; распознавание образов; прогнозная аналитика; имитационное моделирование; пространственный анализ; статистический анализ; визуализация аналитических данных). Big data: применение и возможности. Решения на основе Big data. Рынок Big data в России. Big data в банках. Big data в бизнесе. Big data в маркетинге.

Содержание практических/лабораторных занятий по дисциплине

Тема 1. Методы многомерного статистического анализа и анализа нечисловой информации.

1. 1. Факторный анализ.
- 1.2. Дискриминантный анализ.
- 1.3. Кластерный анализ.
1. 4. Многомерное шкалирование.
1. 5. Методы контроля качества.

Тема 2. Технологии хранения и обработки Больших данных.

- 2.1. Основные направления развития методов обработки и хранения данных.
- 2.2. Проблема хранения неструктурированных данных.
- 2.3. Проблема преобразования данных.
- 2.4. Семантические анализаторы.
- 2.5. Самообучающиеся автоматы.

Тема 3. Программирование обработки и загрузки больших данных.

3.1. Языки для Big Data.

3.2. Фреймворки.

3.3. Базы данных.

3.4. Аналитические платформы.

Тема 4. Аналитика в больших данных.

4.1. Аналитика Big Data — реалии и перспективы в России и мире.

4.2. Технологии и методы анализа, которые используются для анализа Big Data.

4.3. Применение и возможности Big data.

4.4. Решения на основе Big data.

5. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

5.1. Текущий контроль успеваемости

Текущий контроль успеваемости (рейтинг-контроль 1, рейтинг-контроль 2, рейтинг-контроль 3).

Типовые тестовые задания для проведения текущего контроля приведены ниже.

Рейтинг-контроль №1

Тестовые задания:

1. Принятый способ представления данных: показатели должны быть:

- 1) по строкам;
- 2) по ячейкам;
- 3) по столбцам;
- 4) по диагонали.

2. Интервальные данные – это (подчеркните правильные ответы):

- 1) данные с интервалом;
- 2) данные об интервалах;
- 3) количество измерений в каждом интервале;
- 4) количество интервалов в каждом измерении.

3. Среди ниже приведённых нечисловые данные следующие:

- 1) баллы;
- 2) дихотомические;
- 3) ранги;
- 4) рейтинги.

4. Простейшие статистические характеристики – это:

- 1) математическое ожидание;
- 2) среднее;
- 3) с.к.о.;
- 4) дисперсия.

5. Приведение к нормальной форме - это:

- 1) деление на с.к.о.;
- 2) округление;
- 3) деление на среднее;
- 4) деление на константу интегрирования.

6. Какие функции Excel имеют отношение к оцифровке:

- 1) ранг;
- 2) счётесли;
- 3) коррел;
- 4) суммесли.

7. Многомерность в статистике - это:

- 1) переменных больше одной;
- 2) измерений больше 10;
- 3) переменных больше двух;
- 4) измерений больше 5

8. Следующие программы являются специализированными статистическими пакетами:

- 1) EXCEL;
- 2) GRAPHER;
- 3) SPSS;
- 4) STATISTICA.

9. Проверка статистической гипотезы включает в себя:

- 1) ранжирование;
- 2) принятие уровня значимости;
- 3) вычисление эмпирического значения;
- 4) вычисление критического значения.

Рейтинг-контроль №2**Тестовые задания:****1. Кластерный анализ предназначен для:**

- 1) группировки объектов;
- 2) ранжирования объектов;
- 3) группировки показателей;
- 4) ранжирования показателей.

2. Опции кластерного анализа:

- 1) расстояние между группами;
- 2) расстояние между объектами;
- 3) расстояние между показателями;
- 4) расстояние между телами.

3. Кластерный анализ реализован в программах:

- 1) EXCEL;
- 2) SPSS;
- 3) AGRAPHER;
- 4) STATISTICA.

4. Снижение размерности это:

- 1) уменьшение числа измерений;
- 2) уменьшение числа показателей;
- 3) уменьшение числа объектов;
- 4) уменьшение числа знаков.

5. Компонентный анализ реализован в программах:

- 1) EXCEL;

- 2) SPSS;
- 3) AGRAPHER;
- 4) STATISTICA.

6. Методы, относящиеся к снижению размерности:

- 1) факторный анализ;
- 2) регрессия;
- 3) компонентный анализ;
- 4) корреляция.

7. Компонентный анализ позволяет:

- 1) сортировать;
- 2) ранжировать;
- 3) группировать;
- 4) упорядочивать.

8. Дихотомическая шкала это:

- 1) состоящая из “да” и “нет”;
- 2) состоящая из двух чисел;
- 3) состоящая из “истина” и “ложь”;
- 4) состоящая из двух рангов.

9. К нечисловым шкалам относятся:

- 1) номинальная;
- 2) абсолютная;
- 3) интервалов;
- 4) ранговая.

10. Существует шкал для описания данных:

- 1) 4;
- 2) 6;
- 3) 5;
- 4) 7.

1. Количество наблюдений - это:

- 1) размерность;
- 2) ширина;
- 3) объём выборки;
- 4) поверхность выборки.

2. Элементы таблицы сопряжённости называются:

- 1) координаты;
- 2) скорости;
- 3) длины;
- 4) частоты.

3. Методы анализа таблиц сопряжённости:

- 1) Критерий Розенбаума;
- 2) хи-квадрат;
- 3) Критерий Колмогорова-Смирнова;
- 4) критерий Фишера.

4. В ходе анализа таблицы сопряжённости выполняется:

- 1) проверка на соответствие;
- 2) проверка на непротиворечивость;
- 3) проверка на монотонность;
- 4) проверка на значимость.

5. Максимальная размерность таблицы сопряжённости может быть:

- 1) 3;
- 2) 5;
- 3) 10;
- 4) какая угодно.

6. Вычисляемое значение критерия хи-квадрат называется:

- 1) Численное значение;
- 2) реальное значение;
- 3) экспериментальное значение;

4) эмпирическое значение.

7. Вычисляемое значение хи-квадрат сравнивается с:

- 1) критическим значением;
- 2) предельным значением;
- 3) эталонным значением;
- 4) граничным значением.

8. То, с чем сравнивается вычисляемое значение хи-квадрат, вычисляется в EXCEL функцией:

- 1) ХИ2РАСП;
- 2) ХИ2ТЕСТ;
- 3) ХИ2ОБР;
- 4) ХИ2.

9. К коэффициентам связи относятся:

- 1) коэффициент контингенции;
- 2) коэффициент ассоциации;
- 3) Коэффициент Чупрова-Крамера;
- 4) коэффициент коллигации.

10. К разновидности критерия хи-квадрат относятся:

- 1) критерий Вилкоксона;
- 2) информационный критерий;
- 3) критерий Джонкира;
- 4) критерий максимального правдоподобия.

11. Выявление вкладов, вносимых каждой клеткой таблицы, называется:

- 1) разбиение хи-квадрат;
- 2) локализация хи-квадрат;
- 3) анализ хи-квадрат;
- 4) сортировка хи-квадрат.

12 Лог-линейный анализ - это:

- 1) анализ синтеза таблиц;
- 2) анализ достоверности таблиц;

- 3) статистический анализ связи таблиц;
- 4) анализ разброса таблиц.

5.2. Промежуточная аттестация знаний студентов производится по результатам работы в соответствующем семестре, в виде экзамена, который включает в себя ответы на теоретические вопросы.

Вопросы к экзамену

1. Факторный анализ.
2. Дискриминантный анализ.
3. Кластерный анализ.
4. Многомерное шкалирование.
5. Методы контроля качества.
1. Основные направления развития методов обработки и хранения данных.
2. Volume.
3. Закон Мура.
4. Velocity. Variety.
5. Фреймворк Hadoop.
6. Проблема хранения неструктурированных данных.
7. Проблема преобразования данных.
8. Семантические анализаторы.
9. Самообучающиеся автоматы.
10. Языки для Big Data: R.
11. Языки для Big Data: Python.
12. Языки для Big Data: Julia.
13. Языки для Big Data: Java.
14. Языки для Big Data: Scala.
15. Языки для Big Data: MATLAB.
16. Языки для Big Data: Kafka,.
17. Языки для Big Data: Hadoop.
18. Языки для Big Data: Go.
19. Фреймворки для Big Data: Hadoop.
20. Фреймворки для Big Data: Spark.
21. Фреймворки для Big Data: Storm.
22. Базы данных для Big Data: Hive.

23. Базы данных для Big Data: Impala.
24. Базы данных для Big Data: Presto.
25. Базы данных для Big Data: Drill.
26. Аналитические платформы для Big Data: Rapid Miner.
27. Аналитические платформы для Big Data: IBM SPSS Modeler.
28. Аналитические платформы для Big Data: KNIME.
29. Аналитические платформы для Big Data: Qlik Analytics Platform.
30. Аналитические платформы для Big Data: STATISTICA Data Miner.
31. Аналитические платформы для Big Data: Informatica Intelligent Data Platform.
32. Аналитические платформы для Big Data: World Programming System.
33. Аналитические платформы для Big Data: Deductor.
34. Аналитические платформы для Big Data: SAS Enterprise Miner.
35. Zookeeper.
36. Flume.
37. IBM Watson Analytics.
38. Dell EMC Analytic Insights Module.
39. Windows Azure HDInsight.
40. Microsoft Azure Machine Learning.
41. Pentaho Data Integration.
42. Teradata Aster Analytics.
43. SAP BusinessObjects Predictive Analytics.
44. Oracle Big Data Preparation.
45. Аналитика Big Data — реалии и перспективы в России и мире.
46. Data Mining.
47. Краудсорсинг.
48. Смешение и интеграция данных.
49. Машинное обучение.
50. Искусственные нейронные сети.
51. Распознавание образов.
52. Прогнозная аналитика.
53. Имитационное моделирование.
54. Пространственный анализ.
55. Статистический анализ.
56. Визуализация аналитических данных.
57. Big data: применение и возможности.

58. Решения на основе Big data.
59. Рынок Big data в России.
60. Big data в банках.
61. Big data в бизнесе.
62. Big data в маркетинге.

Примерный перечень практических заданий на экзамене

1. Дан набор данных заданной структуры и программа SAS Data step, производящая определенную обработку и вычисления с использованием данного набора. Перепишите эту программу на SAS DS2 с использованием параллельных нитей и созданием пользовательского пакета, чтобы результат обработки сохранился тем же, но код мог выполняться в параллельной среде.

2. Дан набор заданной структуры, постройте модель прогнозирования отклика с использованием процедуры impstat с алгоритмом random forest с заданным числом деревьев. Примените полученную модель к тестовому набору данных той же структуры, визуализируйте полученный график Lift. Постройте на том же наборе модель с использованием высокопроизводительной версии метода GLM. Примените к тестовому набору. Сравните результаты GLM и Random Forest по AUC.

3. Дан текстовый корпус документов, лежащих в указанной директории. Создайте в SAS Text Miner проект, который: выберет файлы с расширением pdf; осуществит парсинг набора с определением частей речи и сохранением в признаковом пространстве только существительных и глаголов; осуществит фильтрацию документов и признаков с использованием заданной схемой определения весов лексем (например, на основе tf-idf); выделит заданное количество ключевых тематик по методу SVD. В ответе укажите топ 5 ключевых слов во второй выявленной тематике. Какой документ имеет наибольший вес в этой тематике?

5.3. Самостоятельная работа обучающегося производится в виде докладов, эссе, презентаций.

Подготовка докладов по следующим темам:

1. Факторный анализ.
2. Дискриминантный анализ.
3. Кластерный анализ.
4. Основные направления развития методов обработки и хранения данных.

5. Проблема хранения неструктурированных данных.
6. Проблема преобразования данных.
7. Семантические анализаторы.
8. Самообучающиеся автоматы.
9. Языки для Big Data.
10. Фреймворки.
11. Базы данных.
12. Аналитические платформы.
13. Аналитика Big Data — реалии и перспективы в России и мире.
14. Big data: применение и возможности.
15. Рынок Big data в России.

Требования по подготовке доклада (эссе).

Эссе - это самостоятельная письменная работа на тему, предложенную преподавателем.

Оно должен содержать:

- введение, содержащее постановку проблемы;
- основную часть, содержащую логически выдержанное изложение темы (предпосылок и путей решения поставленной проблемы);
- краткие выводы, обобщающие позицию автора по проблеме;
- список использованной литературы (указывается только та литература, которой фактически пользовался автор; все случаи использования источников - цитаты, сведения, оценки и т.д. - отмечаются ссылками в виде сносок или примечаний с указанием страниц источника).

Объем эссе должен составлять 7-10 страниц (до 4 тыс. слов) печатного текста (шрифт Times, размер 12, полуторный интервал). Включение в эссе материалов, не имеющих прямого отношения к теме, а также источников, не указанных в базовом списке литературы (в частности, текстов из Интернета), служит основанием для признания работы не соответствующей требованиям или существенного снижения общей оценки.

Эссе оценивается по следующим критериям:

- самостоятельность выполнения работы, способность аргументировано защищать основные положения и выводы. Эссе, выполненное несамостоятельно, по другим критериям не оценивается;
- соответствие формальным требованиям: структура, наличие списка литературы, сносок, грамотность изложения;

- способность сформулировать проблему;
- уровень освоения темы и изложения материала: обоснованность отбора материала, использование первичных источников, способность самостоятельно осмысливать выявленные факты, логика изложения;
- четкость и содержательность выводов.

Тематика эссе

1. Аналитическая деятельность агентства Moody's.
2. Аналитическая продукция Lexis-Nexis.
3. Принцип работы системы Watson.
4. Прогнозы развития систем аналитики по данным ведущих мировых агентств.
5. Направления исследований в области аналитики компании IBM.
6. Технология анализа данных Text Mining.
7. Когнитивная аналитика в продуктах компании IBM
8. Аналитические решения от компании SAS.
9. Современные технологии обработки больших объемов данных.
10. Основные принципы работы OLAP-систем.
11. Перспективы развития машинного перевода
12. Применение систем Business Intelligence.
13. Сравнительный анализ возможностей систем Business Intelligence ведущих производителей (SAP, Oracle, SAS, IBM).
14. Технологии визуализации данных.
15. Геоинформационные системы.

Требования по подготовке презентации

Общие требования к презентации:

- Презентация не должна быть меньше 10 слайдов.
- Первый лист – это титульный лист, на котором обязательно должны быть представлены: название проекта; название выпускающей организации; фамилия, имя, отчество автора; вуз, где учится автор проекта и его группа.
- Следующим слайдом должно быть содержание, где представлены основные части (моменты) презентации. Желательно, чтобы из содержания по гиперссылке можно перейти на необходимую страницу и вернуться вновь на содержание.
- Дизайн-эргономические требования: сочетаемость цветов, ограниченное количество объектов на слайде, цвет текста.

- Презентация не может состоять из сплошного не структурированного текста.
- Последними слайдами урока-презентации должны быть глоссарий и список литературы.

Создание презентации состоит из трех этапов:

I. Планирование презентации – это многошаговая процедура, включающая определение целей, формирование структуры и логики подачи материала. Планирование презентации включает в себя:

1. Определение целей.
2. Определение основной идеи презентации.
3. Подбор дополнительной информации.
4. Планирование выступления.
5. Создание структуры презентации.
6. Проверка логики подачи материала.
7. Подготовка заключения.

II. Разработка презентации – методологические особенности подготовки слайдов презентации, включая вертикальную и горизонтальную логику, содержание и соотношение текстовой и графической информации.

III. Репетиция презентации – это проверка и отладка созданной презентации.

В оформлении презентаций выделяют два блока: оформление слайдов и представление информации на них. Для создания качественной презентации необходимо соблюдать ряд требований, предъявляемых к оформлению данных блоков.

Оформление слайдов:

Стиль	<ul style="list-style-type: none"> · Соблюдайте единый стиль оформления · Избегайте стилей, которые будут отвлекать от самой презентации. · Вспомогательная информация (управляющие кнопки) не должны преобладать над основной информацией (текстом, иллюстрациями).
Фон	<ul style="list-style-type: none"> · Для фона предпочтительны холодные тона
Использование цвета	<ul style="list-style-type: none"> · На одном слайде рекомендуется использовать не более трех цветов: один для фона, один для заголовка, один для текста. · Для фона и текста используйте контрастные цвета. · Обратите внимание на цвет гиперссылок (до и после использования).
Анимационные эффекты	<ul style="list-style-type: none"> · Используйте возможности компьютерной анимации для представления информации на слайде. · Не стоит злоупотреблять различными анимационными эффектами, они не должны отвлекать внимание от содержания информации на слайде.

Представление информации:

Содержание информации	<ul style="list-style-type: none"> · Используйте короткие слова и предложения. · Минимизируйте количество предлогов, наречий, прилагательных. · Заголовки должны привлекать внимание аудитории.
Расположение информации на странице	<ul style="list-style-type: none"> · Предпочтительно горизонтальное расположение информации. · Наиболее важная информация должна располагаться в центре экрана. · Если на слайде располагается картинка, надпись должна располагаться под ней.
Шрифты	<ul style="list-style-type: none"> · Для заголовков – не менее 24. · Для информации не менее 18. · Шрифты без засечек легче читать с большого расстояния. · Нельзя смешивать разные типы шрифтов в одной презентации. · Для выделения информации следует использовать жирный шрифт, курсив или подчеркивание. · Нельзя злоупотреблять прописными буквами (они читаются хуже строчных).
Способы выделения информации	<ul style="list-style-type: none"> · Следует использовать: · рамки; границы, заливку; · штриховку, стрелки; · рисунки, диаграммы, схемы для иллюстрации наиболее важных фактов.
Объем информации	<ul style="list-style-type: none"> · Не стоит заполнять один слайд слишком большим объемом информации: люди могут одновременно запомнить не более трех фактов, выводов, определений. · Наибольшая эффективность достигается тогда, когда ключевые пункты отображаются по одному на каждом отдельном слайде.
Виды слайдов	<p>Для обеспечения разнообразия следует использовать разные виды слайдов:</p> <ul style="list-style-type: none"> · с текстом; · с таблицами; · с диаграммами.

Тематика презентаций

1. Анализ данных социальных сетей.
2. Возможности использования Big Data в различных сферах деятельности.
3. Направления исследований в области аналитики компании IBM.
4. Технология анализа данных Text Mining.
5. Когнитивная аналитика в продуктах компании IBM.
6. Современные технологии обработки больших объемов данных.
7. Основные принципы работы OLAP-систем.
8. Перспективы развития машинного перевода.
9. Краткий обзор российского рынка информационно-аналитических систем.

10. Деятельность ведущих российских информационно-аналитических агентств.
11. Хэш-функции и показатели
12. Методы табличной группировки и консолидации данных
13. Разработка пользовательского приложения на основе базы данных.
14. Применение кластерных алгоритмов в линейных и параллельных вычислениях.
15. Средства реализации методов интеллектуального анализа.

Фонд оценочных материалов (ФОМ) для проведения аттестации уровня сформированности компетенций обучающихся по дисциплине оформляется отдельным документом.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Книгообеспеченность

Наименование литературы: автор, название, вид издания, издательство	Год издания	КНИГООБЕСПЕЧЕННОСТЬ
		Наличие в электронной библиотеке ВлГУ
1	2	3
Основная литература		
1. Анализ данных при помощи Microsoft Power BI и Power Pivot для Excel : практическое руководство / А. Феррари, М. Руссо ; пер. с англ. А. Ю. Гинько. - Москва : ДМК Пресс, 2020. - 288 с. - ISBN 978-5-97060-858-6	2020	https://znanium.com/catalog/document?id=367157
2. Статистические методы анализа данных : учебное пособие / В. Н. Клячкин, Ю. Е. Кувайскова, В. А. Алексева. - Москва : Финансы и Статистика, 2021. - 240 с. - ISBN 978-5-00184-057-2.	2021	https://znanium.com/catalog/product/1831431
3. Данные: хранение и обработка : учебник / Э. Г. Дадян. — Москва : ИНФРА-М, 2021. — 205 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-016447-2	2021	https://znanium.com/catalog/product/1149101
Дополнительная литература		
1 Практический подход к проектированию баз данных : учебное пособие / М. В. Махмутова. - 2-е изд., стер. - Москва : ФЛИНТА, 2023. - 159 с. - ISBN 978-5-9765-3694-4.	2023	https://znanium.com/catalog/product/2091322
2. Базы данных : учебник / Л.И. Шустова, О.В. Тараканов. — Москва : ИНФРА-М, 2023. — 304 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/11549. - ISBN 978-5-16-010485-0..	2023	https://znanium.com/catalog/product/1986697
3. Базы данных: практикум : учебно-	2021	https://znanium.com/catalog/product/

практическое пособие / А. С. Копырин. - Москва : ФЛИНТА, 2021. - 106 с. - ISBN 978- 5-9765-4752-0.		1851992
--	--	-------------------------

6.2. Периодические издания

1. BIG DATA AND ADVANCED ANALYTICS. [HTTP://BIGDATAMINSK.BY](http://BIGDATAMINSK.BY)

6.3. Интернет-ресурсы

1. <http://www.spssbase.com/> Иллюстрированный самоучитель по SPSS

2. <http://www.spss.ru> Официальный сайт российского офиса компании SPSS

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебная аудитория, компьютерный класс с выходом в Internet для проведения занятий лекционного и семинарского типа, текущего и промежуточного контроля, групповых и индивидуальных консультаций.

Демонстрационное оборудование: проектор Panasonic, экран, ноутбук Asus X58Le, 12 станций ThinkCentre M70, звуковые колонки Genius SW-HF5.1, доска настенная.

Количество посадочных мест: 18.

Расположена по адресу: 600005, Российская Федерация, Владимирская область, г.о. город Владимир, г. Владимир, ул. Горького, д. 79, 2 этаж учебного корпуса № 6, 52,5 м², № 2.

Перечень используемого лицензионного программного обеспечения: операционная система семейства Microsoft Windows, пакет офисных программ Microsoft Office, справочная правовая система «Консультант Плюс» (инсталлированный ресурс ВлГУ).

Примечание

В соответствии с нормативно-правовыми актами для инвалидов и лиц с ограниченными возможностями здоровья при необходимости тестирование может быть проведено только в письменной или устной форме, а также могут быть использованы другие материалы контроля качества знаний, предусмотренные рабочей программой дисциплины.

Рабочую программу составил: ст. преподаватель Виноградов Д.В.

Рецензент (представитель работодателя):

Начальник отдела информатизации арбитражного суда Владимирской области Дигилин Д.Е.

Программа рассмотрена и одобрена на заседании кафедры БИиЭ

протокол № 1 от «30» августа 2023 года.

Заведующий кафедрой: д.э.н., профессор Тесленко И.Б.

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии
направления 01.03.05 Статистика

протокол № 1 от «05» сентября 2023 года.

Председатель комиссии: к.э.н., доцент Ярьес О.Б.