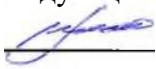



ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФАКУЛЬТЕТ ФИЛОСОФИИ И СОЦИОЛОГИИ

Утверждено:
на заседании кафедры социологии и
работы с молодежью
протокол № 9 от «22» мая 2020 г.
Заведующий кафедрой

/Шайхисламов Р.Б.

Согласовано:
Председатель УМК факультета
философии и социологии

 /Хабибуллина З.Н.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

дисциплина
Анализ больших данных в социологии

Часть, формируемая участниками образовательных отношений

программа магистратуры

Направление подготовки (специальность)
39.04.01 Социология

Направленность (профиль) подготовки
«Социальное проектирование в сферах коммуникаций и маркетинга»

Квалификация
магистр

Разработчик (составитель)
Доцент, к.э.н., доцент



В. М. Тимирьянова

Для приема: 2020 г.

Уфа 2020 г.

Составитель: Тимирьянова В. М.

Рабочая программа дисциплины утверждена на заседании кафедры социологии и работы с молодежью протокол от «22» мая 2020 г. № 9

Заведующий кафедрой



/ Шайхисламов Р.Б.

Список документов и материалов

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций	4
2. Цель и место дисциплины в структуре образовательной программы	5
3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)	5
4. Фонд оценочных средств по дисциплине	5
4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.	5
4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.	7
5. Учебно-методическое и информационное обеспечение дисциплины	21
5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	21
5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы	21
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине	22
Приложение 1	23

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций

По итогам освоения дисциплины обучающийся должен достичь следующих результатов обучения:

Категория (группа) компетенций¹ (при наличии ОПК)	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
<i>Системное и критическое мышление</i>	<i>ПК-1. Способен к разработке предложений по совершенствованию методов проведения социологических и маркетинговых исследований</i>	<i>ПК-1.1. Знать методы проведения социологических и маркетинговых исследований использующие технологии больших данных.</i>	<i>Знать технологии больших данных. Знать направления использования больших данных</i>
		<i>ПК-1.2. Уметь совершенствовать методы проведения социологических и маркетинговых исследований в результате включения технологий больших данных.</i>	<i>Уметь использовать технологии больших данных</i>
		<i>ПК-1.3. Владеть навыками разработки предложений по совершенствованию методов проведения социологических и маркетинговых исследований в результате включения технологий больших данных</i>	<i>Владеть технологиями больших данных</i>
	<i>ПК-2. Способен разрабатывать теоретико-методологические основы социологического исследования с позиций различных социологических</i>	<i>ПК-2.1. Знать теоретико-методологические основы социологического исследования и возможности применения в них технологии больших данных.</i>	<i>Знать технологии больших данных. Знать направления использования больших данных</i>

¹ Указывается только для УК и ОПК (при наличии).

	<i>теорий и парадигм</i>	<i>ПК-2.2. Уметь развивать теорию и методологию исследования в результате включения технологий больших данных.</i>	<i>Уметь использовать технологии больших данных</i>
		<i>ПК-2.3. Владеть навыками разработки теоретико-методологические основы исследования с позиций технологий больших данных</i>	<i>Владеть технологиями больших данных Владеть методиками выявления тенденций по результатам обработки больших данных.</i>

2. Цель и место дисциплины в структуре образовательной программы

Дисциплина «Анализ больших данных в социологии» относится к вариативной части. Дисциплина изучается на 1 курсе во 2 семестре.

Цели изучения дисциплины: формирование знаний, навыков и умений, необходимых для применения технологии больших данных при проведении социологических и маркетинговых исследований.

3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)

Содержание рабочей программы представлено в Приложении № 1.

4. Фонд оценочных средств по дисциплине

4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотношенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения	
		«Не зачтено»	«Зачтено»
ПК-1.1. Знать методы проведения социологических и маркетинговых исследований использующие технологии больших данных.	Знать технологии больших данных. Знать направления использования больших данных	Отсутствуют или фрагментарные представления о технологиях больших данных и направлениях использования больших данных	Сформированные систематические представления о технологиях больших данных и направлениях использования больших данных

<p>ПК-1.2. Уметь совершенствовать методы проведения социологических и маркетинговых исследований в результате включения технологий больших данных.</p>	<p>Уметь использовать технологии больших данных</p>	<p>Отсутствуют или фрагментарные умения использовать технологии больших данных</p>	<p>Сформированное умение использовать технологии больших данных</p>
<p>ПК-1.3. Владеть навыками разработки предложений по совершенствованию методов проведения социологических и маркетинговых исследований в результате включения технологий больших данных</p>	<p>Владеть технологиями больших данных</p>	<p>Фрагментарное владение технологиями больших данных</p>	<p>Успешное и систематическое применение технологий больших данных</p>
<p>ПК-2.1. Знать теоретико-методологические основы социологического исследования и возможности применения в них технологии больших данных.</p>	<p>Знать технологии больших данных. Знать направления использования больших данных</p>	<p>Отсутствуют или фрагментарные представления о технологиях больших данных и направлениях использования больших данных</p>	<p>Сформированные систематические представления о технологиях больших данных и направлениях использования больших данных</p>
<p>ПК-2.2. Уметь развивать теорию и методологию исследования в результате включения технологий больших данных.</p>	<p>Уметь использовать технологии больших данных</p>	<p>Отсутствуют или фрагментарные умения использовать технологии больших данных</p>	<p>Сформированное умение использовать технологии больших данных</p>
<p>ПК-2.3. Владеть навыками разработки теоретико-методологические основы</p>	<p>Владеть технологиями больших данных Владеть методиками</p>	<p>Фрагментарное владение технологиями больших данных</p>	<p>Успешное и систематическое применение технологий больших данных</p>

социологического исследования с позиций технологий больших данных	выявления тенденций по результатам обработки больших данных.		
---	--	--	--

4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные средства
ПК-1.1. Знать методы проведения социологических и маркетинговых исследований использующие технологии больших данных.	Знать технологии больших данных. Знать направления использования больших данных	Вопросы к зачету, тестирование, письменная контрольная работа, публикация статей, аудиторная работа - дискуссия, устный опрос, индивидуальное задание
ПК-1.2. Уметь совершенствовать методы проведения социологических и маркетинговых исследований в результате включения технологий больших данных.	Уметь использовать технологии больших данных	Вопросы к зачету, тестирование, письменная контрольная работа, публикация статей, аудиторная работа - дискуссия, устный опрос, индивидуальное задание
ПК-1.3. Владеть навыками разработки предложений по совершенствованию методов проведения социологических и маркетинговых исследований в результате включения технологий больших данных	Владеть технологиями больших данных	Вопросы к зачету, письменная контрольная работа, публикация статей, устный опрос, индивидуальное задание
ПК-2.1. Знать теоретико-методологические основы социологического исследования и возможности применения в них технологии	Знать технологии больших данных. Знать направления использования больших данных	Вопросы к зачету, тестирование, письменная контрольная работа, публикация статей, аудиторная работа - дискуссия, устный опрос, индивидуальное задание

больших данных.		
ПК-2.2. Уметь развивать теорию и методологию исследования в результате включения технологий больших данных.	Уметь использовать технологии больших данных	Вопросы к зачету, тестирование, письменная контрольная работа, публикация статей, аудиторная работа - дискуссия, устный опрос, индивидуальное задание
ПК-2.3. Владеть навыками разработки теоретико-методологические основы социологического исследования с позиций технологий больших данных	Владеть технологиями больших данных Владеть методиками выявления тенденций по результатам обработки больших данных.	Вопросы к зачету, письменная контрольная работа, публикация статей, устный опрос, индивидуальное задание

Показатели сформированности компетенции для очной формы обучения:

Критериями оценивания при *модульно-рейтинговой системе* являются баллы, которые выставляются преподавателем за виды деятельности (оценочные средства) по итогам изучения модулей (разделов дисциплины), перечисленных в рейтинг-плане дисциплины *для зачета*: текущий контроль – максимум 50 баллов; рубежный контроль – максимум 50 баллов, поощрительные баллы – максимум 10).

Шкалы оценивания:

зачтено – от 60 до 110 рейтинговых баллов (включая 10 поощрительных баллов),
не зачтено – от 0 до 59 рейтинговых баллов).

Показатели сформированности компетенций для заочной формы обучения:

Критериями оценивания являются оценки, которые выставляются преподавателем за виды деятельности (оценочные средства) по итогам выполнения всех заданий и прохождения аттестации по дисциплине.

Шкалы оценивания:

«зачтено»;
«не зачтено».

Система оценивания для заочного обучения

Оценка	Критерий	Индикатор (показатель)
«зачтено»	усвоение программы в полном объеме	задание выполнено без замечаний, полное и логически стройное изложение содержания при ответе или в отчете, тесное увязывание теории вопроса с практикой, отсутствие затруднений с объяснением всех аспектов выполнения задания, хорошее владение умениями и навыками по программе, знание монографической литературы, наличие умений самостоятельно обобщать и излагать материал
«не зачтено»	невладение значительной (и значимой) частью материала программы	задание не выполнено, допуск обучающимся при ответе принципиальных ошибок, большие затруднения при выполнении практических работ, ответ свидетельствует об отсутствии знаний по предмету

Система оценивания для очного обучения

Система оценивания для очного обучения по видам заданий приведена после каждого задания.

Вопросы для самоконтроля по дисциплине

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Какие главные характеристики Big Data?
4. Какие данные занимают больше мировой памяти относительно остальных?
5. Какие понятия содержит в себе принцип трех "V"?
6. Чем анализ больших данных отличается от традиционного анализа?
7. Каковы причины появления технологий больших данных?
8. Какие примеры использования на разных уровнях управления существуют?
9. Какие положительные эффекты использования технологий больших данных могут быть получены в социологии, в сферах коммуникаций и маркетинга?
10. Какие проблемы применения больших данных в социологии, в сферах коммуникаций и маркетинга отмечаются?
11. В чем суть национальной программы «Цифровая экономика»?
12. Какими компетенциями должен обладать работник в условиях цифровой трансформации?
13. Какие источники получения больших данных существуют?
14. Какие виды и форматы данных существуют?
15. В чем суть структуризации и каталогизации?
16. Какими принципами стоит руководствоваться при работе с данными?
17. Какие технологии обработки данных существуют?
18. Что включает экосистема Hadoop?
19. Какие возможности и ограничения использования ресурсов среды программирования R при анализе больших данных вы знаете?
20. В чем суть вычислительной модели MapReduce?
21. В чем суть HDFS?
22. Что является целью кластеризации?
23. Что такое Apache Hadoop?
24. В чем преимущества решений на базе Hadoop?
25. Что такое MapReduce?

26. Какими достоинствами и недостатками обладает MapReduce?
27. Какому основному принципу следует HDFS?

Вопросы к зачету

1. Сущность больших данных и перспективы их использования в социологии, в сферах коммуникаций и маркетинга
2. Условия и возможности использования больших данных в социологии, в сферах коммуникаций и маркетинга
3. Примеры использования больших данных в социологии, в сферах коммуникаций и маркетинга в зарубежной практике
4. Примеры использования больших данных в социологии, в сферах коммуникаций и маркетинга на национальном уровне
5. Примеры использования больших данных в социологии, в сферах коммуникаций и маркетинга на региональном уровне
6. Примеры использования больших данных в социологии, в сферах коммуникаций и маркетинга на частном уровне
7. Возможные эффект использования больших данных в других областях жизнедеятельности.
8. Основные цели обработки Big Data в социологии, в сферах коммуникаций и маркетинга?
9. Главные характеристики Big Data.
10. Причины появления технологий больших данных.
11. Проблемы применения больших данных в социологии, в сферах коммуникаций и маркетинга.
12. Национальная программ «Цифровая экономика»: сущность и перспективы
13. Компетенции работника в условиях цифровой трансформации.
14. Источники получения больших данных.
15. Виды и форматы данных.
16. Распараллеливание данных: сущность и алгоритмы.
17. Экосистема Hadoop.
18. Возможности и ограничения использования ресурсов среды программирования R при анализе больших данных.
19. Вычислительная модель MapReduce.
20. В чем суть HDFS?

Оценка	Критерий	Индикатор (показатель)
«зачтено»	усвоение программы в полном объеме	задание выполнено без замечаний, полное и логически стройное изложение содержания при ответе или в отчете, тесное увязывание теории вопроса с практикой, отсутствие затруднений с объяснением всех аспектов выполнения задания, хорошее владение умениями и навыками по программе, знание монографической литературы, наличие умений самостоятельно обобщать и излагать материал
«не зачтено»	невладение значительной (и значимой) частью материала программы	задание не выполнено, допуск обучающимся при ответе принципиальных ошибок, большие затруднения при выполнении практических работ, ответ свидетельствует об отсутствии знаний по предмету

Планы семинарских занятий

Модуль 1. Сущность и источники больших данных в социологии, в сферах коммуникаций и маркетинга

1. Сущность технологий больших данных и примеры использования в социологии, в сферах коммуникаций и маркетинга

- понятие Big data;
- причины появления технологий больших данных;
- примеры использования в социологии, в сферах коммуникаций и маркетинга;
- положительные эффекты использования технологий больших данных;
- проблемы применения больших данных в социологии, в сферах коммуникаций и маркетинга;
- национальная программа «Цифровая экономика»;
- модель компетенций работника в условиях цифровой трансформации;
- автоматизированные системы в социологии, в сферах коммуникаций и маркетинга;

2. Источники, объемы и скорость накопления больших данных

- источники получения больших данных: открытые и закрытые;
- объемы информации и места хранения;
- виды данных: потоковые и пакетные;
- форматы данных;
- структурированные и неструктурированные данные;
- структуризация данных и каталогизация;
- скорость накопления данных в социологии, в сферах коммуникаций и маркетинга;
- принципы работы с данными;

Модуль 2. Технологии анализа больших данных и их практическая реализация в социологии, в сферах коммуникаций и маркетинга

1. Существующие технологии и принципы обработки больших данных

- существующие технологии обработки данных;
- Bigdata-ориентированные информационные системы в социологии, в сферах коммуникаций и маркетинга;
- возможности и ограничения использования ресурсов среды программирования R при анализе больших данных;
- оптимизация структуры объектов данных в ходе вычислений;
- распараллеливание вычислений в R;
- пакет RНadoop;
- система Hadoop и R;
- вычислительная модель Map/Reduce;
- распределенная файловая система HDFS;
- моделирование в условиях ограниченного объема памяти;

2. Основные прикладные задачи для больших данных

- возможности применения технологий больших данных в социологии, в сферах коммуникаций и маркетинга;
- алгоритм проведения анализа на основе больших данных;
- технология работы с большими данными;

3. Проведение анализа с применением технологии больших данных, их интерпретация и формирование аналитических отчетов

- работа в системе Hadoop;
- подходы к формированию отчета по результатам анализа;

- виды представления результатов анализа.

Аудиторная работа проводится в форме семинаров – устного опроса, дискуссий, выполнения индивидуальных заданий; тестирования, написания контрольной работы.

Вопросы для семинаров в рамках аудиторной работы (устный опрос):

Модуль 1. Сущность и источники больших данных в социологии, в сферах коммуникаций и маркетинга

1. Сущность технологий больших данных и примеры использования в социологии, в сферах коммуникаций и маркетинга

Проработать и обосновать основные вопросы темы:

- 1) Понятие и причины появления технологий больших данных.
- 2) Использование больших данных на национальном, региональном и муниципальном уровнях управления.
- 3) Эффекты использования технологий больших данных.
- 4) Автоматизированные системы в социологии, в сферах коммуникаций и маркетинга.

2. Источники, объемы и скорость накопления больших данных

Проработать и обосновать основные вопросы темы:

- 1) Открытые/закрытые, потоковые/пакетные, структурированные/неструктурированные данные.
- 2) Структуризация, каталогизация, распараллеливание.
- 3) Принципы работы с большими данными.

Модуль 2. Технологии анализа больших данных и их практическая реализация в социологии, в сферах коммуникаций и маркетинга

1. Существующие технологии и принципы обработки больших данных

Проработать и обосновать основные вопросы темы:

- 1) Принципы обработки больших данных.
- 2) Возможности и ограничения использования различных пакетов и систем при анализе больших данных;
- 3) Возможности моделирования в условиях ограниченного объема памяти.

2. Основные прикладные задачи для больших данных

Проработать и обосновать основные вопросы темы:

- 1) Основные направления применения технологий больших данных в социологии, в сферах коммуникаций и маркетинга.
- 2) Алгоритм проведения анализа на основе больших данных.

3. Проведение анализа с применением технологии больших данных, их интерпретация и формирование аналитических отчетов

Проработать и обосновать основные вопросы темы:

- 1) Визуализация данных.
- 2) Аналитические отчеты в социологии, в сферах коммуникаций и маркетинга.

Критерии оценивания	Количество баллов
студент подготовил полную информацию и организовал хорошее выступление	2
Студент не подготовил выступление, но владеет материалом и активно участвует в обсуждении	1
Студент не подготовил выступление, не участвовал в обсуждении	0

Дискуссия

Дискуссия проводится по вопросу: Каковы перспективы применения больших данных в социологии, в сферах коммуникаций и маркетинга?

Критерии оценивания	Количество баллов
Студент подготовил 2-3 вопроса, активно участвовал в дискуссии, грамотно ответив не менее чем на 3 вопроса, продемонстрировал уважение ко всем участникам дискуссии и проявил понимание всех обсуждаемых вопросов.	7-9
Студент подготовил 1-2 вопроса, активно участвовал в дискуссии, грамотно ответив не менее чем на 1 вопрос; продемонстрировал уважение ко всем участникам дискуссии	4-6
Студент не сумел подготовить ни один вопрос, но активно участвовал в дискуссии, грамотно ответив не менее чем на 3 вопроса; продемонстрировал уважение ко всем участникам дискуссии.	1-3
Студент не подготовил ни один вопрос и не участвовал в дискуссии; продемонстрировал неуважение ко всем участникам дискуссии, проявил полное непонимание всех обсуждаемых вопросов	0

Индивидуальное задание

Варианты заданий

Задание 1.

MapReduce – это модель распределенной обработки данных, предложенная компанией Google для обработки больших объёмов данных на компьютерных кластерах для экосистемы Hadoop. Механизм ее работы предлагается изучить путем пошаговой реализации ее алгоритмов на условном примере.

Задача студента - выполнить подсчет количества разнородных элементов в наборе данных, выданных преподавателем: в следующей последовательности, согласно алгоритма MapReduce:

Шаг 1: Сохраните набор данных в 4 разделах (как в HDFS). Цель научиться балансировать нагрузку на ресурсы.

Шаг 2: Сопоставьте данные. Цель понять суть кластеризации и присвоения ключей.

Шаг 3: Сортировка и перемешивание. Цель научиться балансировать нагрузку на ресурсы.

Шаг 4: Сгенерировать пары "ключ-значение".

Задание 2.

Hadoop Distributed File System (HDFS) – распределённая файловая система, позволяющая хранить информацию практически неограниченного объёма.

Cloudera Virtual Machine (VM) Image - преднастроенный виртуальный образ/машина создаваемый на персональном компьютере.

Задача студента - посчитать, сколько раз встречается слово в тексте.

Шаг 1: Загрузите и установите VirtualBox.

Шаг 2: Загрузите и установите образ виртуальной машины (VM) Cloudera.

Шаг 3: Запустите Cloudera VM.

Шаг 4: Загрузите файл в HDFS с текстом, выданный преподавателем;

Шаг 5: Запустите приложение WordCount.

Шаг 6: Скопируйте результаты WordCount из Hadoop Distributed File System (HDFS).

Критерии оценивания	Количество баллов
Задание выполнено полностью: цель выполнения задания успешно достигнута; основные понятия определены; работа выполнена в полном объёме.	8-10
Задание выполнено полностью, но не очень качественно: цель выполнения задания успешно достигнута; основные понятия определены; работа выполнена в полном объёме, работа выполнена небрежно, ссылки не открываются.	7-8
Задание выполнено: цель выполнения задания достигнута; наличие правильных обоснованных ответов; однако работа выполнена не в полном объёме.	1-6
Задание не выполнено, цель выполнения задания не достигнута.	0

Тестирование

Тестирование проводится на основе имеющихся тестовых заданий или на основе вопросов-тестов, подготовленных студентами.

1) **Тестирование на основе** имеющихся тестовых заданий из фонда оценочных средств

Тестовые задания

1. BigData – это ...

- (1) Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации, или обработки.
- (2) Комплексный набор методов обработки структурированных и неструктурированных данных колоссальных объемов.
- (3) Колоссальный объем данных, собранных человечеством.
- (4) Класс в Java, предназначенный для хранения данных от 100 Гб

2. Объем накопленных человечеством цифровых данных на 2012 год измеряется:

- (1) петабайтами
- (2) зеттабайтами
- (3) экзабайтами
- (4) йоттабайтами

3. Укажите фактор, способствовавший появлению тренда больших данных

- (1) маркетинговые кампании крупных корпораций
- (2) снижение издержек на хранение данных
- (3) появление новых технологий обработки потоковых данных
- (4) выпуск баз данных с обработкой данных в памяти

4. Какие вероятные разочарования тренда больших данных?

- (1) из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных
- (2) из-за угрозы безопасности личной жизни (privacy) граждан будут упрощены процедуры сбора данных, что приведёт к падению ценности больших данных
- (3) нет

5. Отметьте значимые события, повлиявшие на формирование тренда больших данных:

- (1) разработка Hadoop
- (2) изобретение принципа MapReduce
- (3) разработка языка Python

6. Какие данные занимают больше мировой памяти относительно остальных?

- (1) Structured Data
- (2) Unstructured Data
- (3) Semi-Structured Data
- (4) Quasi-Structured Data

7. Выберите верный ответ

- (1) большие данные – это обработка или хранение более 1 Тб информации
- (2) проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
- (3) большие данные – это огромная PR-акция крупных вендоров и не более того
- (4) большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект

8. Выберите неверный ответ:

- (1) большие данные – это данные объёма свыше 1 Тб
- (2) проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
- (3) большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров
- (4) большие данные как правило не структурированы

9. Отметьте те из вариантов, в которых данные структурированы:

- (1) данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word
- (2) таблица с ежедневными показаниями температуры помещения за год в файле формата csv
- (3) текст педагогической поэмы А.С. Макаренко, представленный в формате PDF
- (4) библиотека фильмов, представленных в формате mpeg4 на одном жестком диске

10. Перечислите четыре основных характеристики Big Data:

- (1) Virtualization, Volume, Variability, Vehicle
- (2) Variety, Velocity, Volume, Value
- (3) Verification, Volume, Velocity, Visualization
- (4) Video, Value, Variety, Volume

11. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:

- (1) Горизонтальное масштабирование
- (2) Вертикальное масштабирование
- (3) Master- slave репликация
- (4) Peer-to-peer репликация

12. Принцип MapReduce состоит в том, чтобы

- (1) производить вычисления на узлах, где информация изначально была сохранена
- (2) использовать вычислительные мощности систем хранения

(3) использовать функциональное программирование для решения задач массивно-параллельной обработки

13. Что из этого является недостатком MapReduce?

- (1) Фиксированный алгоритм обработки данных
- (2) Масштабируемость
- (3) Отказоустойчивость
- (4) Возможность автоматического распараллеливания

14. Данные имеющие определенный тип, формат и структуру (например, транзакции) являются:

- (1) Структурированными
- (2) Полуструктурированными
- (3) Квазиструктурированными
- (4) Неструктурированными

15. Какая компания создала технологию MapReduce?

- (1) Google
- (2) Yahoo
- (3) EMC
- (4) Oracle

16. Выберите одно неверное высказывание про MapReduce:

- (1) интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изначально была сохранена
- (2) MapReduce – это две операции: распределения и сборки данных
- (3) MapReduce был придуман разработчиками Hadoop
- (4) MapReduce был анонсирован разработчиками Google

17. Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?

- (1) 100Гб
- (2) 1Тб
- (3) 100Тб
- (4) 1Пб

18. Человек покупает товары через интернет. Государство хочет знать насколько могут возрасти такие продажи в ближайшем будущем и когда. К какому типу относится эта задача анализа данных?

- (1) прогнозирование
- (2) кластеризация
- (3) классификация
- (4) цензурирование

19. Инвестиционный фонд интересуется тем, почему часть финансируемых им проектов успешно переходят на второй год, а часть - нет. К какому типу относится эта задача анализа данных?

- (1) поиск информативных признаков
- (2) построение решающего правила
- (3) классификация
- (4) цензурирование

20. Инвестиционный фонд имеет ряд проектов, который успешно переходят на второй год финансирования и тех, кто не переходит. Как бы в данном случае формулировалась задача поиска информативных признаков?

- (1) определить, почему ряд проектов успешно переходят на второй год, а ряд – нет
- (2) определить для нового проекта, перейдёт ли он через год на второй этап финансирования или нет
- (3) восстановить некоторые характеристики проектов, которые изначально не заполнялись
- (4) определить критерий успешности

21. Инвестиционный фонд имеет ряд проектов, который успешно переходят на второй год финансирования и тех, кто не переходит. Фонд поставил задачу определить критерий успешности проекта. К какому типу задач анализа данных наиболее близка эта задача?

- (1) прогнозирование
- (2) построение решающего правила
- (3) поиск информативных признаков
- (4) цензурирование

22. Поликлиникой ставится цель определения структуры своих клиентов с точки зрения числа обращений. К какому типу относится эта задача анализа данных?

- (1) прогнозирование
- (2) кластеризация
- (3) классификация
- (4) цензурирование

23. Поликлиника обладает некоторыми данными о клиентах и о их возрасте. Как бы в данном случае формулировалась задача кластеризации?

- (1) определить основные группы клиентов
- (2) определить, сколько раз придет тот или иной клиент в следующем периоде
- (3) определить, когда вернётся тот или иной клиент

24. Компания, проводящая социологические опросы, испытывает сложности с верификацией данных, поступающих от волонтеров непосредственно опрашиваемых респондентов: многие анкеты заполнены не полностью; волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет. К какому типу наиболее близка эта задача анализа данных?

- (1) прогнозирование
- (2) кластеризация
- (3) классификация
- (4) цензурирование

25. Компания, проводящая социологические опросы получает анкеты от волонтеров, непосредственно опрашиваемых респондентов. При каких условиях разумна постановка задачи цензурирования?

- (1) многие анкеты заполнены не полностью
- (2) стало известно, что волонтеры фальсифицируют результаты опроса, самостоятельно заполняя часть анкет
- (3) от заказчика поступило требование уничтожить часть анкет, содержащих информацию о руководителях страны
- (4) Часть анкет пришла в негодность, что не позволяет считать информацию с них со 100% уверенностью

26. Компания, проводящая социологические опросы, испытывает сложности с верификацией данных, поступающих от волонтеров непосредственно опрашиваемых респондентов: многие анкеты заполнены не полностью; волонтеры фальсифицируют результаты опроса,

самостоятельно заполняя часть анкет. К какому типу задач анализа данных здесь прибегать не придётся?

- (1) заполнение пробелов
- (2) цензурирование
- (3) прогнозирование
- (4) классификация

27. С некоторой периодичностью на госпредприятии списываются группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Руководство заинтересовано в сокращении количества проверок, при сохранении точности выявления ошибочного списания на уровне 97%. Требуется выявлять сомнительные акты списания, подлежащие обязательной проверке аудитором. К какому типу относится эта задача анализа данных?

- (1) прогнозирование
- (2) кластеризация
- (3) классификация
- (4) цензурирование

28. С некоторой периодичностью на госпредприятии списываются группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Как бы в данном случае формулировалась задача классификации?

- (1) определить характерные признаки ошибочных списаний
- (2) научиться автоматически выявлять ошибочные списания с ожидаемой ошибкой не ниже 97%
- (3) классифицировать типичные ошибки и составить их список
- (4) определить три категории: "ошибочные", "под сомнением", "безошибочные" и найти правило отнесения к этим категориям

29. С некоторой периодичностью на госпредприятии списываются группы расходных материалов на различных участках учета. Для выявления ошибок, акты списания выборочно проверяются аудитором. Определены три категории: "ошибочные", "под сомнением", "безошибочные". К какому типу задач анализа данных относится задача о построении правила автоматического отнесения списаний к этим категориям.

- (1) поиск информативных признаков
- (2) кластеризация
- (3) классификация
- (4) цензурирование

30. К какому типу шкал относится шкала "очень плохо"-"плохо"-"средне"-"хорошо"-"очень хорошо"?

- (1) порядковая
- (2) абсолютная
- (3) бинарная
- (4) номинальная

Критерии оценивания	Количество баллов
90-100% верных ответов	5
75-90% верных ответов	4
60-75% верных ответов	3

40-60% верных ответов	2
20-40% верных ответов	1
Верных ответов менее 20%	0

2) **Тестирование на основе** вопросов-тестов, составленных студентами, после чего осуществляется взаимная проверка-тестирование.

Студенты готовят вопросы-тесты с использованием изученного в ходе лекций и практических занятий материала, лист с заданиями автором подписывается. Затем студенты меняются заданиями и отвечают на вопросы-тесты своих одногруппников, возвращают лист с заданиями авторам вопросов. Авторы вопросов проверяют правильность ответов и отмечают верные и неверные ответы.

Критерии оценивания	Количество баллов
Студент подготовил 3-5 вопросов-тестов и успешно ответил на все 3-5 вопросов-тестов одногруппников	4-5
Студент подготовил 3-5 вопросов-тестов и успешно ответил на 1-2 вопросов-тестов одногруппников	2-3
Студент подготовил 2-3 вопроса-теста, но не сумел успешно ответить на 1 вопрос-тест одногруппников	1
Студент не подготовил и не ответил ни на один вопрос-тест одногруппников	0

Письменная контрольная работа

Студент самостоятельно и без применения лекций, интернета и прочих источников отвечает на два из вопросов ниже письменно. Работа подписывается и сдается.

Примерные вопросы для письменных работ

1. В чем заключается сущность технологий больших данных
2. Приведите примеры использования больших данных в социологии, в сферах коммуникаций и маркетинга
3. Опишите основные источники больших данных в социологии, в сферах коммуникаций и маркетинга
4. Опишите принципы и алгоритм работы с большими данными.
5. Опишите существующие технологии обработки больших данных
6. Обозначьте возможности и ограничения использования различных пакетов и систем при анализе больших данных.
7. Предложите прикладную задачу для сфер коммуникаций и маркетинга предусматривающую применение технологий больших данных.
8. С помощью каких средств могут быть визуализированы данные и в какой форме может быть представлен отчет по результатам анализа больших данных.

Критерии оценивания	Количество баллов
Ответ полный по обоим вопросам, по существу, грамотно изложен (без грамматических, орфографических и прочих ошибок)	13-15
Суть вопросов отражена, но есть незначительные ошибки или неточности	9-12
Суть вопросов отражена, но есть ошибки (в фактах, датах, названиях, в ответе есть непонимание терминов)	5-8
Ответ содержит ошибки и не отражает сути вопросов	2-4

Студент не написал ответ ни на один из двух вопросов или пользовался шпаргалками	0
--	---

Подготовка статьи

Описание статьи:

Подготовка статьи является одним из творческих заданий, выполняемых студентами.

В случае ее публикации студент получает поощрительные 10 баллов. В случае, если статья не была опубликована, она учитывается как одно из творческих заданий.

Структура статьи: аннотация, актуальность вопроса, формулировка проблемы, опыт по решению проблемы и проработанность вопроса (авторы, публикации), предложенный вариант решения проблемы, заключение, список использованных источников.

По содержанию статья написана грамотно, отражает владение терминологией, умение работать с источниками информации и самостоятельно делать выводы. Автор демонстрирует критическое мышление и предлагает конкретное управленческое решение, мероприятие для решения проанализированного вопроса.

Презентация по структуре аналогична статье, сделана в формате Power Point и отражает личный вклад автора в решение вопроса.

Описание методики оценивания:

Выдержана рекомендуемая структура статьи, презентация подготовлена и представлена в группе студентов (на конференции), в статье заметен личный вклад автора, мысли сформулированы четко и логично.

Критерии оценивания	Количество баллов
Статья написана на основе проработки не менее 5 источников, логично изложена и представлена ясно и четко, есть однозначно хорошее владение материалом (даны ответы на вопросы группы)	8-10
Статья написана на основе проработки не менее 3 источников, представлена ясно и четко, есть однозначно хорошее владение материалом (даны ответы на вопросы группы)	5-7
Статья написана на основе проработки не менее 3 источников, нет однозначно хорошего владения материалом (нет ответов на вопросы группы)	1-4
Статья не подготовлена	0

5. Учебно-методическое и информационное обеспечение дисциплины

5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Келлехер, Д. Наука о данных: базовый курс : [16+] / Д. Келлехер, Б. Тирни ; науч. ред. З. Мамедьяров ; пер. с англ. М. Белоголового. – Москва : Альпина Паблишер, 2020. – 224 с. : схем., табл. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=598235> (дата обращения: 26.08.2020). – ISBN 978-5-9614-3170-4.

2. Хусаинова, Г. Я. Базы данных [Электронный ресурс]: учеб.-метод. пособие / Г. Я. Хусаинова; МОиН РФ; СФ БашГУ; Под ред. И. Г. Хусаинова, Е. М. Карасева. — Стерлитамак: Изд-во СФ БашГУ, 2017 — 79 с. — Электрон. версия печ. публикации. — Доступ возможен через Электронную библиотеку БашГУ. — <URL:https://elib.bashedu.ru/dl/corp/Husainova_G_Vazy_dannyh_ump_2017.pdf>

Дополнительная литература:

3. Провалов, В.С. Информационные технологии управления : учебное пособие / В.С. Провалов. – 4-е изд., стер. – Москва : Флинта, 2018. – 374 с. – (Экономика и управление). – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=69111>.

4. Цифровизация экономики: мир, Россия, регионы : монография / И.В. Митрофанова, И.А. Рябова, О.В. Фетисова и др. – Москва ; Берлин : Директ-Медиа, 2019. – 73 с. : ил., схем., табл. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=570917> (дата обращения: 26.08.2020). – Библиогр.: с. 64-69. – ISBN 978-5-4499-0439-3. – DOI 10.23681/570917.

5. Адлер, Ю. Практическое руководство по статистическому управлению процессами : практическое руководство : [12+] / Ю. Адлер, В.Л. Шпер ; ред. В. Ионов. – Москва : Альпина Паблишер, 2019. – 234 с. : ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=570307> (дата обращения: 26.08.2020). – Библиогр. в кн. – ISBN 978-5-9614-2053-1.

6. Косых, Е. С. Делопроизводство в органах государственной и муниципальной власти [Электронный ресурс]: учебно-методическое пособие для магистрантов юридического факультета / Е. С. Косых; СФ БашГУ; отв. ред. Г. А. Иванцова. — Стерлитамак: Изд-во СФ БашГУ, 2019 — 50 с. — Электрон. версия печ. публикации. — Доступ возможен через Электронную библиотеку БашГУ. — <URL:https://elib.bashedu.ru/dl/read/Kosyh_Deloproizvodstvo_up_2019.pdf>.

5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы

№	Наименование Интернет-ресурса	Ссылка (URL) на Интернет ресурс
1.	Федеральная служба государственной статистики	www.gks.ru
2.	Национальная программа «Цифровая экономика 2024»	digital.ac.gov.ru/

Перечень программного обеспечения и информационных справочных систем:

– Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г.

Лицензии – бессрочные.

– Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии – бессрочные.

– Система централизованного тестирования БашГУ (Moodle). GNU General Public License Version 3, 29 June 2007

– КонсультантПлюс. Договор № 28826 от 09.01.2019 г. Лицензии бессрочные.

– ЭБС «Университетская библиотека онлайн»;

– ЭБС издательства «Лань»;

– ЭБС «Электронный читальный зал»;

– БД периодических изданий на платформе EastView: «Вестники Московского университета», «Издания по общественным и гуманитарным наукам»;

– Научная электронная библиотека;

– БД диссертаций Российской государственной библиотеки.

Также доступны следующие зарубежные научные ресурсы баз данных:

– Web of Science;

– Scopus;

– Издательство «Taylor&Francis»;

– Издательство «Annual Reviews»;

– «Computers & Applied Sciences Complete» (CASC) компании «EBSCO»

– Архивы научных журналов на платформе НЭИКОН (Cambridge University Press, SAGE Publications, Oxford University Press);

– Информационная система «Единое окно доступа к образовательным ресурсам» (<http://window.edu.ru>).

- Install VirtualBox. Лицензия открытая. <https://www.virtualbox.org/wiki/Downloads>

- Cloudera VM Лицензия открытая. https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.4.2-0-virtualbox.zip.

6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

<i>Наименование специализированных аудиторий, кабинетов, лабораторий</i>	<i>Вид занятий</i>	<i>Наименование оборудования, программного обеспечения</i>
<i>1</i>	<i>2</i>	<i>3</i>
<i>Аудитория</i>	<i>Лекции</i>	<i>Компьютер, мультимедийный проектор, экран, доска и т.д.</i>
<i>Лаборатория</i>	<i>Лабораторные работы</i>	<i>Наименование оборудования (при необходимости) (например, прибор, установка, набор и т.д.)</i>
<i>Компьютерный класс</i>	<i>Практические занятия</i>	<i>Компьютеры, имеющие информационно-вычислительные аналитические системы, которые включают в себя базы данных, методы обработки информации для ...</i>

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФАКУЛЬТЕТ ФИЛОСОФИИ И СОЦИОЛОГИИ

СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ

дисциплины Анализ больших данных в социологии на 2 семестр

Очная

форма обучения

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (з.е. / часов)	2/72
Учебных часов на контактную работу с преподавателем:	
лекций	18
практических/ семинарских	18
лабораторных	
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	0,2
Учебных часов на самостоятельную работу обучающихся (СР)	35,8
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	

Форма(ы) контроля:
зачет 2 семестр

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)				Основная и дополнительная литература, рекомендуемая студентам (номера из списка)	Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		ЛК	ПР/СЕМ	ЛР	СРС			
1	2	3	4	5	6	7	8	9
	Модуль 1. Сущность и источники больших данных в социологии, в сферах коммуникаций и маркетинга							
1.	Сущность технологий больших данных и примеры использования в социологии, в сферах коммуникаций и маркетинга	2	2		4	1, 3, 4, 5	Вопросы для самоконтроля; подготовка к дискуссии	Дискуссия, тестирование
2.	Источники, объемы и скорость накопления больших данных	4	4		4	1, 2, 4, 6	Вопросы для самоконтроля;	Тестирование Контрольная работа
	Модуль 2. Технологии анализа больших данных и их практическая реализация в социологии, в сферах коммуникаций и маркетинга							
3.	Существующие технологии и принципы обработки больших данных	4	4		9	1, 2, 3	Вопросы для самоконтроля;	Индивидуальное задание
4.	Основные прикладные задачи для больших данных	4	4		9	1, 3	Вопросы для самоконтроля; Доработка индивидуального задания,	Индивидуальное задание
7.	Проведение анализа с применением технологии больших данных, их интерпретация и формирование аналитических отчетов	4	4		9,8	1, 5	Подготовка материала для презентации отчета	Презентация отчета
	Всего часов:	18	18		35,8			

