



Государственное бюджетное образовательное учреждение высшего образования
Московской области

ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ
Проректор по учебной работе
Е.К. Самаров
« 25 » 2021г.



**ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И
ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И УПРАВЛЯЮЩИХ
СИСТЕМ**

**РАБОЧАЯ ПРОГРАММА
ДИСЦИПЛИНЫ
«АНАЛИЗ БОЛЬШИХ ДАННЫХ»**

Направление подготовки: 01.03.02. Прикладная математика и информатика

Профиль: Искусственный интеллект и управление в ракетно-космических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная

Королев 2021

Автор: к.т.н., доц. Аббасова Т.С. Рабочая программа дисциплины: Анализ больших данных: МГОТУ, 2021

Рецензент: к.т.н., доц. Исаева Г.Н.

Рабочая программа составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего профессионального образования (ФГОС ВО) по направлению подготовки бакалавров 09.03.02 «Информационные системы и технологии» и Учебного плана, утвержденного Ученым советом МГОТУ.

Протокол № 13 от 22.06, 2021 года.

Рабочая программа рассмотрена и одобрена на заседании кафедры:

Заведующий кафедрой (ФИО, ученая степень, звание, подпись)	Артюшенко В.М. д.т.н. профессор 	Артюшенко В.М. д.т.н., профессор 	Артюшенко В.М. д.т.н., профессор 	
Год утверждения (переутверждения)	2021	2022	2023	2024
Номер и дата протокола заседания кафедры	№15 от 02.06.21	№13 от 03.06.22	№14 от 03.05.23	

Рабочая программа согласована:

Руководитель ОПОП _____



к.т.н., доц. И.В. Бугай

Рабочая программа рекомендована на заседании УМС:

Год утверждения (переутверждения)	2021	2022	2023	2024
Номер и дата протокола заседания УМС	№7 от 15.06.21	№5 от 21.06.22	№6 от 16.05.23	

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОПОП

Целью изучения дисциплины «Анализ больших данных» является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу бакалавриата, для решения различных задач практической и научно-исследовательской деятельности.

Дисциплина направлена на частичное формирование следующих компетенций:

общефессиональные компетенции (ОПК):

- Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности (ОПК-3);

профессиональные компетенции (ПК):

- Способность использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ (ПК-3).

Основными задачами дисциплины являются:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- статистический, инженерный, интуитивный анализ информации;
- приобретение практических навыков работы с программой Power BI.

После завершения освоения данной дисциплины студент должен:

Знать

- Методики поиска, сбора и обработки больших данных; актуальные российские и зарубежные источники информации в сфере профессиональной деятельности; метод системного анализа.
- Математические методы анализа больших данных.
- Алгоритмы обработки больших данных.

Уметь

- Применять методики поиска, сбора и обработки информации; осуществлять критический анализ и синтез информации, полученной из разных источников; применять системный подход для решения поставленных задач.
- Использовать и адаптировать существующие математические методы и системы программирования для анализа больших данных
- Использовать современные методы разработки и реализации алгоритмов обработки больших данных.

Владеть

- Методами поиска, сбора и обработки, критического анализа и синтеза информации; методикой системного подхода для решения поставленных задач.
- Методами и способами анализа больших данных.
- Методами и способами обработки больших данных.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина «Анализ больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений, основной профессиональной образовательной программы подготовки бакалавров по направлению подготовки 01.03.02. «Прикладная математика и информатика».

Дисциплина базируется на ранее изученных дисциплинах: «Теория вероятностей и математическая статистика», «Технологии и среды программирования», «Математическая логика и теория алгоритмов: ОПК-1, ОПК-3, ОПК-4, ПК-2».

Знания и компетенции, полученные при освоении дисциплины, являются базовыми при выполнении выпускной квалификационной работы бакалавра.

3. Объем дисциплины и виды учебной работы

Общая трудоемкость дисциплины составляет 2 зачетных единиц, 72 часа.

Таблица 1

Виды занятий	Всего часов	Семестр	Семестр	Семестр	Семестр
		пятый	шестой	седьмой	восьмой
Общая трудоемкость	72	72			
ОЧНАЯ ФОРМА ОБУЧЕНИЯ					
Аудиторные занятия	32	32			
Лекции (Л)	16	16			
Практические занятия (ПЗ)	16	16			
Лабораторные работы (ЛР)	-	-			
Самостоятельная работа	40	40			
Курсовые работы (проекты)	-	-			
Расчетно-графические работы	-	-			
Контрольная работа, домашнее задание	+	+			
	-	-			
Текущий контроль знаний	Тест	+			
Вид итогового контроля	зачет	зачет			
ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ: ОТСУТСТВУЕТ					

4. Содержание дисциплины

4.1. Темы дисциплины и виды занятий

Таблица 2

Наименование разделов и тем	Лекции, час. очн/заочн, час	Практические занятия, очн/заочн, час	Занятия в интерактивной форме очн/заочн час	Код компетенций
Тема 1. Введение в анализ больших данных	4/-	4/-	2/-	ОПК-3
Тема 2. Технологии добычи, хранения и обработки больших данных	4/-	4/-	2/-	ПК-3
Тема 3. Методы анализа и визуализации данных	4/-	4/-	2/-	ПК-3
Тема 4. Программные сервисы для анализа больших данных	4/-	4/-	2/-	ПК-3
Итого по первому семестру:	16/-	16/-	8/-	

4.2. Содержание тем дисциплины

Тема 1. Введение в анализ больших данных

Обзор источников информации. Основные определения, термины, задачи анализа больших данных. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных. Обзор современных популярных программных средства анализа данных: Statistica, SPSS, Excel, R-Studio, Power-BI и другие. Загрузка, анализ и преобразование больших данных в различных программных средах.

Тема 2. Технологии добычи, хранения и обработки больших данных

Алгоритмы для решения задач добычи данных и очень больших данных. Технология MapReduce – важное средство распараллеливания алгоритмов. Алгоритмы хэширования с учетом близости и потоковой обработки данных для тщательного анализа больших данных. Показатель PageRank. Технологии хранения больших данных (базы данных, хранилища данных, многомерные хранилища данных). Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных. Обзор средств бизнес-аналитики больших данных. Экосистема Microsoft Power BI. Обзор источников данных для Power BI и особенностей работы с каждым из них.

Тема 3. Методы анализа и визуализации больших данных

Методы статистического анализа данных. Нахождение частых предметных наборов и кластеризация. Метод анализа drill down/up, при котором производится последовательное увеличение уровня подробности рассматриваемых данных с целью выявления причин отклонений или неудовлетворительных результатов. Понижение размерности данных. Визуализация аналитических данных – представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации как для получения результатов, так и для использования в качестве исходных данных для дальнейшего анализа.

Тема 4. Программные сервисы для анализа больших данных

Набор программных сервисов Microsoft, которые работают вместе, превращая несвязанные источники данных компании в целостные интерактивные отчеты. Базы данных, файлы Excel, данные из облачных источников и интернета, текстовые файлы как источники данных. Инфографика в интерактивных сервисах. Программные средства поддержки. Улучшение возможностей работы с цифровыми технологиями обработки больших данных. Power BI для выполнения задач самостоятельной и корпоративной бизнес-аналитики.

5. Перечень учебно-методического обеспечения для самостоятельной работы по дисциплине

1. Методические указания для обучающихся по освоению дисциплины.

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Структура фонда оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведена в Приложении 1.

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

Основная литература:

1. Кравченко, Ю.А. Тенденции развития компьютерных технологий: учебное пособие / Ю.А. Кравченко, Э.В. Кулиев, Д.В. Заруба; Южный федеральный университет, Инженерно-технологическая академия. – Таганрог: Южный федеральный университет, 2017. – 107 с.: схем., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=493214> (дата обращения: 29.09.2020). – Библиогр. в кн. – ISBN 978-5-9275-2360-3. – Текст: электронный.

2. Лесковец, Ю. Анализ больших наборов данных / Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман; пер. с англ. А.А.Слинкина. - Москва:

ДМК Пресс, 2016. - 498 с. - ISBN 978-5-97060-190-7. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1027845> (дата обращения: 06.10.2020). – Режим доступа: по подписке.

Дополнительная литература:

1. Анализ больших данных и приложение [Электронный ресурс] / Haitao Xiong .— : Aviation Industry Press, 2015 .— 326 с. — ISBN 978-7-5165-0956-2 .— Режим доступа: <https://lib.rucont.ru/efd/717979>

2. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - Москва: Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3. - Текст: электронный. - URL: <https://znanium.com/catalog/product/551044> (дата обращения: 06.10.2020). – Режим доступа: по подписке.

Рекомендуемая литература:

1. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013. – 464 с.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <http://www.rusedu.info/> Сайт «Информационные технологии в образовании»
2. <http://nit.miem.edu.ru/> Сайт «Международная студенческая школа-семинар «Новые информационные технологии»».

9. Методические указания для обучающихся по освоению дисциплины

Методические указания для обучающихся по освоению дисциплины (модуля) приведены в Приложении 2.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения: MSOffice (для создания отчетов), MS Excel, Power BI.

Информационные справочные системы: не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды МГОТУ:

Рабочая программа и методическое обеспечение по дисциплине «Анализ больших данных».

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Лекционные занятия:

- аудитория, оснащенная презентационной техникой (проектор, экран);
- комплект электронных презентаций / слайдов на темы:
 - 1) Задачи и методы анализа больших данных
 - 2) Методы и алгоритмы решения задач добычи больших данных
 - 3) Методы и алгоритмы анализа добычи больших данных
 - 4) Технологии хранения и анализа больших данных
 - 5) Технологии OLAP
 - 6) Технологии Data Mining
 - 7) Методы статистического анализа данных
 - 8) Сервисы Power BI

Практические занятия:

- учебный класс, оснащенный вычислительной техникой (ПК);
- рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
- рабочие места студентов, оснащенные компьютерами с доступом в Интернет.

**ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И
ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
И УПРАВЛЯЮЩИХ СИСТЕМ**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО
ДИСЦИПЛИНЕ
« АНАЛИЗ БОЛЬШИХ ДАННЫХ »
(Приложение 1 к рабочей программе)**

Направление подготовки: 01.03.02. Прикладная математика и информатика

Профиль: Искусственный интеллект и управление в ракетно-космических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная

Королев 2021

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

№ п/п	Индекс компетенции	Содержание компетенции (или ее части)*	Раздел дисциплины, обеспечивающий формирование компетенции (или ее части)	В результате изучения раздела дисциплины, обеспечивающего формирование компетенции (или ее части), обучающийся должен:		
				знать	уметь	владеть
1	ОПК-3	Способен применять и модифицировать математические модели для решения задач в области профессиональной деятельности	Тема 1.	Методики поиска, сбора и обработки информации; актуальные российские и зарубежные источники информации в сфере профессиональной деятельности; метод системного анализа	Применять методики поиска, сбора и обработки информации; осуществлять критический анализ и синтез информации, полученной из разных источников; применять системный подход для решения поставленных задач	Методами поиска, сбора и обработки, критического анализа и синтеза информации; методикой системного подхода для решения поставленных задач.
2	ПК-3	Способен выполнять работы по созданию (модификации) и сопровождению информационных систем	Тема 2. Тема 3. Тема 4.	Математические методы анализа больших данных. Алгоритмы обработки больших данных	Использовать и адаптировать существующие математические методы и системы программирования для анализа больших данных Использовать современные методы разработки и реализации алгоритмов обработки больших данных	Методами и способами анализа и обработки больших данных

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Код компетенции	Инструменты, оценивающие сформированность компетенции	Показатель оценивания компетенции	Критерии оценки
ОПК-3	Доклад в форме презентации	<p>А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов</p> <p>Б) частично сформирована:</p> <ul style="list-style-type: none"> •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; <p>В) не сформирована (компетенция не освоена) – 2 и менее баллов</p>	<p>Проводится устно с использованием мультимедийных систем, а также с использованием технических средств</p> <p>Время, отведенное на процедуру – 10 - 15 мин.</p> <p>Неявка – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> 1.Соответствие представленной презентации заявленной тематике (1 балл). 2.Качество источников и их количество при подготовке доклада и разработке презентации (1 балл). 3.Владение информацией и способность отвечать на вопросы аудитории (1 балл). 4.Качество самой представленной презентации (1 балл). 5.Оригинальность подхода и всестороннее раскрытие выбранной тематики (1 балл). <p>Максимальная сумма баллов - 5 баллов.</p> <p>Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.</p>
ПК-3	Контрольная работа	<p>А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов</p> <p>Б) частично сформирована:</p> <ul style="list-style-type: none"> •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; <p>В) не сформирована (компетенция не</p>	<ol style="list-style-type: none"> 1. Проводится в форме письменной работы 2.Время, отведенное на процедуру – семестр. <p>Неявка на защиту контрольной работы – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> 1.Соответствие содержания контрольной работы заявленной тематике (1 балл). 2.Качество источников и их количество при подготовке работы (1 балл). 3.Владение информацией и способность отвечать на вопросы аудитории (1 балл).

		освоена) – 2 и менее баллов	4.Качество самой представленной работы (1 балл). 5. Использование специализированного программного обеспечения (1 балл). 6.Оригинальность подхода и всестороннее раскрытие выбранной тематики (1 балл). Максимальная сумма баллов - 6 баллов. Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.
ОПК-2	Письменное задание	А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов Б) частично сформирована: •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; В) не сформирована (компетенция не освоена) – 2 и менее баллов	1. Проводится в форме письменной работы 2.Время, отведенное на процедуру – 10 - 15 мин. Неявка – 0. Критерии оценки: 1.Соответствие ответа заявленной тематике (0-5 баллов). Максимальная сумма баллов - 5 баллов. Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.
ПК-3	Реферат	А) полностью сформирована (компетенция освоена на высоком уровне) – 5 баллов Б) частично сформирована: •компетенция освоена на продвинутом уровне – 4 балла; •компетенция освоена на базовом уровне – 3 балла; В) не сформирована (компетенция не освоена) – 2 и менее баллов	Проводится в письменной форме Критерии оценки: 1.Соответствие содержания реферата заявленной тематике (1 балл). 2.Качество источников и их количество при подготовке работы (1 балл). 3.Владение информацией и способность отвечать на вопросы аудитории (1 балл). 4.Качество самой представленной работы (1 балл). 5.Оригинальность подхода и всестороннее раскрытие выбранной тематики (1 балл). Максимальная сумма баллов - 5 баллов. Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для

			текущего контроля. Оценка проставляется в электронный журнал.
--	--	--	---

3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Тематика докладов в презентационной форме

1. Дополненная аналитика: практический подход к автоматизации внутренней аналитики
2. Формулировки задач при использовании технологий Data Mining.
3. Построение модели и кластеризация данных с неожиданными характеристиками.
4. Цифровой помощник для маневрового диспетчера.
5. Предсказание цен на сырье и конечные продукты на основе машинного обучения и продвинутой аналитики
6. Способы добычи больших данных.
7. Технологии обработки больших данных.
8. Статистический анализ больших данных.
9. Методы анализа неструктурированных данных текстов.
10. Методы анализа неструктурированных данных сетей.
11. Методы анализа неструктурированных данных изображений.
12. Имитационное моделирование сложных социальных систем.
13. Проблемы машинного обучения.
14. Предиктивные модели.
15. Платформа геоаналитики.
16. Big data на службе семейного бюджета.
17. Скоринговые модели.
18. Учет новых слов в языковых моделях классификации.
19. «Прикладные» нейросети в браузере и на десктопе клиента.
20. Разработка скрининговых тестов.
21. Технологии анализа финансовых потоков и прогнозов.
22. Внедрение управления на основе данных в нефтегазовой отрасли.
23. Искусственный интеллект для снижения вредных выбросов в атмосферу на производстве.
24. Построение системы realtime-аналитики.
25. Границы применимости при создании аналитических решений.

Тематика рефератов

1. Средства анализа данных в офисных приложениях с помощью OLAP-технологий.

2. Методы Data Mining для обучения ассоциативным правилам (англ. association rule learning).
3. Методы классификации в Data Mining.
4. Методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным в Data Mining.
5. Кластерный анализ в Data Mining.
6. Регрессионный анализ в Data Mining.
7. Техники смещения и интеграции разнородных данных для возможности глубинного анализа.
8. Цифровая обработка сигналов для интеграции разнородных данных.
9. Обработка естественного языка для интеграции разнородных данных.
10. Машинное обучение с учителем и без учителя,
11. Использование моделей, построенных на базе статистического анализа или машинного обучения, для получения комплексных прогнозов на основе базовых моделей.
12. Искусственные нейронные сети для анализа больших данных.
13. Прогнозная аналитика на космическом производстве.
14. Имитационное моделирование больших данных.
15. Пространственный анализ с использованием топологической информации в данных.
16. Пространственный анализ с использованием геометрической информации в данных.
17. Пространственный анализ с использованием географической информации в данных.
18. Статистический анализ больших данных с помощью A/B-тестирования.
19. Статистический анализ больших данных с помощью временных рядов.
20. Адаптивные и мультипликативные методы прогнозирования.
21. Дисперсионный анализ влияния качественных факторов.
22. Применение программного обеспечения для решения прикладных задач прогнозирования.
23. Большие данные в информационных технологиях.
24. Подходы, инструменты и методы обработки структурированных данных огромных объёмов.
25. Подходы, инструменты и методы обработки неструктурированных данных огромных объёмов.

Тематика письменных заданий

1. Перечислите методы многомерного анализа больших данных.
2. Опишите принципы факторного анализа.
3. Опишите принципы кластерного анализа.
4. Опишите принципы таксономии.
5. Опишите принципы распознавания образов.

6. Перечислите достоинства факторного анализа.
7. Перечислите достоинства кластерного анализа.
8. Перечислите достоинства таксономии.
9. Перечислите достоинства распознавания образов.
10. Перечислите недостатки факторного анализа.
11. Перечислите недостатки кластерного анализа.
12. Перечислите недостатки таксономии.
13. Перечислите недостатки распознавания образов.
14. Сформулируйте задачу кластеризации.
15. Опишите алгоритм последовательной кластеризации.
16. Опишите метод представления матрицы расстояний с помощью дендограммы.
17. Дайте характеристику Big Data на мировом рынке.
18. Охарактеризуйте Big Data в России.
19. Определите понятие Data Mining.
20. Определите понятие OLAP.
21. В чем состоит когнитивный анализ данных?
22. Какие модели данных вы знаете?
23. Опишите основную идею корреляционного анализа.
24. Опишите основную идею регрессионного анализа.
25. Опишите основную идею дисперсионного анализа.

4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Формой контроля знаний по дисциплине «Анализ больших данных» являются две текущие аттестации в виде тестов и итоговая аттестация в форме зачета.

Неделя текущего контроля	Вид оценочного средства	Код компетенций, оцениваемых знаний, умения, навыки	Содержание оценочного средства	Требования к выполнению	Срок сдачи (неделя семестра)	Критерии оценки по содержанию и качеству с указанием баллов
Согласно графика учебного процесса	Тестирование 1,2	ОПК-2	25 вопросов	Компьютерное тестирование ; время отведенное на процедуру -30 минут	Результаты тестирования предоставляются в день проведения процедуры	Критерии оценки определяются процентным соотношением. Не явка -0 Удовлетворительно - от 51% правильных ответов. Хорошо - от 70%. Отлично – от 90%.
Согласно графика учебного процесса	зачет	ПК-3	2 практических задания	Зачет проводится в письменной форме, путем ответа на вопрос и решения практического задания. время отведенное на процедуру – 0,25 часа на студента.	Результаты предоставляются в день проведения зачета	Критерии оценки: «Зачтено»: знание основных понятий предмета; умение использовать и применять полученные знания на практике; работа на семинарских занятиях; знание основных научных теорий, изучаемых предметов; ответ на вопросы билета. «Не зачтено»: демонстрирует частичные знания по темам дисциплин; незнание основных понятий предмета; неумение использовать и применять

						полученные знания на практике; не работал на семинарских занятиях; не отвечает на вопросы.
--	--	--	--	--	--	--

4.1. Типовые вопросы, выносимые на тестирование

Тесты используются в режиме промежуточного контроля. По форме заданий выбраны закрытые тесты (с выборочным ответом). Каждому вопросу соответствует один вариант ответа.

1) Объём накопленных человечеством цифровых данных на 2020 год измеряется:

- (?) петабайтами
- (!) зеттабайтами
- (?) эксабайтами
- (?) йоттабайтами

2) В одном байте:

- (!) 8 бит
- (?) 6 бит
- (?) 4 бита
- (?) 16 бит

3) В одном килобайте:

- (!) 1024 байт
- (?) 1000 байт
- (?) 1012 байт
- (?) 1048 байт

4) В одном килобите:

- (?) 1024 бит
- (!) 1000 бит
- (?) 1012 бит
- (?) 1048 бит

5) Информация – это:

- (!) нематериальная сущность, при помощи которой с любой точностью можно описывать реальные (материальные), виртуальные (возможные) и понятийные сущности.
- (?) материальная сущность, при помощи которой с любой точностью можно описывать реальные (материальные), виртуальные (возможные) и понятийные сущности.
- (?) материальная сущность, при помощи которой с заданной точностью можно описывать реальные (материальные), виртуальные (возможные) и понятийные сущности.

(?) нематериальный объект, при помощи, которого с заданной точностью можно описывать реальные (материальные), виртуальные (возможные) и понятийные сущности.

6. Хранение информации

(!) процесс передачи информации во времени, связанный с обеспечением неизменности состояний материального носителя

(?) перенос сведений от одного момента времени к другому

(?) любые действия со сведениями, выполняемые по детерминированному алгоритму, которые приводят к изменению их вида, а также содержательности, ценности или полезности

(?) перенос сведений из одной точки пространства в другую.

7) Сколько Петабайт в Зеттабайте?

(?) 64 байт

(?) 1012 байт

(?) 1048 байт

(!) 1024 байт

8) укажите фактор, способствовавший появлению тренда больших данных

(?) разработка новых программ для визуализации данных

(!) снижение издержек на хранение данных

(?) появление новых технологий обработки потоковых данных

(?) выпуск баз данных с обработкой данных в памяти

8) Отметьте значимое событие, повлиявшее на формирование тренда больших данных:

(?) разработка операционной системы Linux

(!) изобретение принципа MapReduce

(?) разработка языка Python

(?) победа DeepBlue в матче с Г.Каспаровым

9) Какое предложение верно?

(?) большие данные – это обработка или хранение более 1 Тб информации

(?) большие данные – это огромная PR-акция крупных вендоров и не более того

(!) проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна

(?) большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект

10) Отметьте вариант, в котором данные структурированы:

(?) данные о продажах компании, представленные в виде ежемесячных отчетов в формате MS Word

(!) таблица с ежедневными показаниями температуры помещения за год в файле формата csv

(?) текст педагогической поэмы А.С. Макаренки, представленный в формате PDF

(?) библиотека фильмов, представленных в формате mp4 на одном жестком диске

4.2. Типовые вопросы, выносимые на зачет

1. Источники больших данных.
2. Основные определения, термины, задачи анализа больших данных.
3. Понятие Data Mining.
4. Когнитивный анализ данных.
5. Методики сбора данных.
6. Алгоритмы для решения задач добычи данных и очень больших данных.
7. Распараллеливание алгоритмов при обработке больших данных
8. Задачи технологии MapReduce.
9. Алгоритмы хэширования с учетом близости и потоковой обработки данных для тщательного анализа больших данных.
10. Показатель PageRank при обработке больших данных.
11. Принципы обработки больших данных в базах данных.
12. Принципы обработки больших данных в хранилищах данных.
13. Принципы обработки больших данных в многомерных хранилищах данных.
14. Первичная обработка больших данных.
15. Методы статистического анализа данных. Deskриптивная статистика, параметрические, непараметрические, номинальные методы.
16. Корреляционный анализ.
17. Регрессионный анализ.
18. Дисперсионный анализ.
19. Кластерный анализ.
20. Дискриминантный анализ.
21. Факторный анализ.
22. Метод анализа больших данных drill down/up.
23. Классификация данных. Методы решения задачи классификации данных. Примеры классификации данных.
24. Кластеризация данных. Методы решения задачи кластеризации данных. Примеры кластеризации данных.
25. Методы прогнозирования данных при обработке информации.
26. Статистика и отчеты. Визуализация справочной информации. Программные и аппаратные средства поддержки.
27. Визуализация аналитических данных – представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации. Программные средства поддержки.
28. Инфографика в интерактивных сервисах. Программные средства поддержки.
29. Улучшение возможностей работы с цифровыми технологиями обработки больших данных.

30. Power BI для выполнения задач самостоятельной и корпоративной бизнес-аналитики.

***ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ
И ТЕХНОЛОГИЙ***

***КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
И УПРАВЛЯЮЩИХ СИСТЕМ***

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО
ОСВОЕНИЮ ДИСЦИПЛИНЫ
«АНАЛИЗ БОЛЬШИХ ДАННЫХ»
(Приложение 2 к рабочей программе)**

Направление подготовки: 01.03.02. Прикладная математика и информатика

Профиль: Искусственный интеллект и управление в ракетно-космических системах

Уровень высшего образования: бакалавриат

Форма обучения: очная

1. Общие положения

Цель дисциплины: теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу бакалавриата, для решения различных задач практической и научно-исследовательской деятельности.

Задачи дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- статистический, инженерный, интуитивный анализ информации;
- приобретение практических навыков работы с программой Power BI.

2. Указания по проведению практических занятий

Тема 1. Введение в анализ больших данных

Практическое занятие 1

Статистический анализ данных обработки экспериментов

Вид практического занятия: компьютерное моделирование

Образовательные технологии: технологии компьютерного обучения.

Обработка статистических данных с помощью встроенных функций MS Excel; применение Пакета анализа и его инструментов: «Генерация случайных чисел», «Гистограмма», «Описательная статистика» на примере обработки измерений скорости движения.

Продолжительность практического занятия 2/- часа.

Практическое занятие 2

Регрессионный и корреляционный анализ больших данных

Вид практического занятия: компьютерное моделирование

Образовательные технологии: технологии компьютерного обучения.

Методы регрессионного и корреляционного анализа и результаты объединения этих методов с помощью программы MS EXCEL. Прогнозирование, планирование, развитие приоритетных направлений, принятие управленческих решений.

Продолжительность практического занятия 2/- часа.

Тема 2. Технологии добычи, хранения и обработки больших данных

Практическое занятие 3.

Изучение интерфейса и возможностей Power BI

Вид практического занятия: компьютерное моделирование.

Образовательные технологии: технологии компьютерного обучения.

Регистрация в Power BI и системные требования. Знакомство с инструментами Power BI и системой четырёх этапов решения аналитических задач. Работа с Power BI Service. Фильтры и интерактивность. Фокусировка, сортировка, подсказки.

Продолжительность практического занятия 2/- часа.

Практическое занятие 4

Загрузка и преобразование больших данных

Вид практического занятия: компьютерное моделирование.

Образовательные технологии: технологии компьютерного обучения.

Подготовка исходных данных к дальнейшему анализу. Работа со сложными запросами и функциями. Экспорт данных. Создание отчетов, копий и их кастомизация. Интерфейс Power BI Desktop. Источники данных для Power BI. Подготовка данных к работе. Импорт данных из файлов Excel, CSV. Обработка, преобразования и очистка данных.

Продолжительность практического занятия 2/- часа.

Тема 3. Методы анализа и визуализации данных

Практическое занятие 5

Визуализация больших данных

Вид практического занятия: практическая работа в группах.

Образовательные технологии: технологии компьютерного обучения.

Выбор и настройка визуальных элементов в зависимости от цели визуализации и вида данных. Типы графиков и возможности форматирования. Принципы выбора диаграмм. Основные ошибки при визуализации данных. Условное форматирование и форматирование таблиц. Настройка взаимодействия между визуальными элементами. Настройка фильтров и управляющих элементов.

Продолжительность практического занятия 2/- часа.

Практическое занятие 6

Публикация данных и совместная работа с отчётами

Вид практического занятия: практическая работа в группах.

Образовательные технологии: технологии компьютерного обучения.

Переход от Power BI Desktop в сервис Power BI для совместной работы. Отличия дашбордов и отчётов. Составление и работа с дашбордами. Машинное обучение в скрытой аналитике Power BI. Power BI в браузере и Excel.

Продолжительность практического занятия 2/- часа.

Тема 4. Программные сервисы для анализа больших данных

Практическое занятие 7

Power BI Service и Power BI Mobile

Вид практического занятия: практическая работа в группах.

Образовательные технологии: технологии компьютерного обучения.

Публикация отчёта на портал powerbi.com. Создание и настройка дашбордов для разных ролей. Совместная работа с контентом. Автоматическое обновление данных и Power BI Gateway. Мобильное приложение Power BI.

Продолжительность практического занятия 2/- часа.

Практическое занятие 8

Безопасность данных в Power BI Service

Вид практического занятия: практическая работа в группах.

Образовательные технологии: технологии компьютерного обучения.

Безопасность данных и разграничение доступа. Группы и настройка доступов. Настройка оповещений.

Продолжительность практического занятия 2/- часа.

3. Указания по проведению лабораторного практикума

Не предусмотрен учебным планом.

4. Указания по проведению самостоятельной работы студентов

№ п/п	Наименование блока (раздела) дисциплины	Виды СРС
1.	Тема 1. Ведение в анализ больших данных	1. Способ представления больших данных с помощью диаграммы разброса 2. Способ представления больших данных с помощью лепестковой диаграммы 3. Способ представления больших данных с помощью тепловой диаграммы 4. Способ представления больших данных с помощью облака тегов
2.	Тема 2. Технологии добычи, хранения и обработки больших данных	1. Обработка полевых сейсморазведочных данных в модификациях 2D. 2. Обработка полевых сейсморазведочных данных в модификациях 3D. 3. Обработка данных в системах дистанционного зондирования земли.

3	Тема 3. Методы анализа и визуализации больших данных	<ol style="list-style-type: none"> 1. Корреляционный анализ. 2. Регрессионный анализ. 3. Дисперсионный анализ. 4. Кластерный анализ. 5. Дискриминантный анализ. 6. Факторный анализ.
4	Тема 4. Программные сервисы для анализа больших данных	<ol style="list-style-type: none"> 1. Средства информационного описания в сети Интернет 2. Информационно-поисковые каталоги 3. Мета-поисковые системы 4. Оценка показателей эффективности поисковых систем

5. Указания по проведению контрольных работ

5.1. Требования к структуре

Структура контрольной работы должна способствовать раскрытию темы: иметь титульный лист, содержание, введение, основную часть, заключение, список литературы.

5.2. Требования к содержанию

5.2.1. Во введении обосновывается актуальность темы, определяется цель работы, задачи и методы исследования.

5.2.2. При определении целей и задач исследования необходимо правильно их формулировать. Так, в качестве цели не следует употреблять глагол «сделать». Правильно будет использовать глаголы: «раскрыть», «определить», «установить», «показать», «выявить» и т.д.

5.2.3. Основная часть работы включает вопросы, каждый из которых посвящается решению задач, сформулированных во введении, и заканчивается констатацией итогов.

5.2.4. Приветствуется иллюстрация содержания работы таблицами, графическим материалом (рисунками, схемами и т.п.).

5.2.5. Необходимо давать ссылки на используемую Вами литературу.

5.2.6. Заключение должно содержать сделанные автором работы выводы, итоги исследования.

5.2.7. Вслед за заключением идет список литературы, который должен быть составлен в соответствии с установленными требованиями. Если в работе имеются приложения, они оформляются на отдельных листах, и должны быть соответственно пронумерованы.

5.3. Требования к оформлению.

Объём контрольной работы – 5...10 страниц формата А4, напечатанного с одной стороны текста (1,5 интервал, шрифт Times New Roman). Контрольная работа должна быть также представлена в электронном виде.

Тематика контрольных работ

1. Использование специализированных расчетных пакетов для анализа и обработки многомерных данных.
2. Использование технологий OLAP для анализа и обработки больших данных.
3. Использование технологий Data Mining для анализа и обработки больших данных.
4. Использование CALS-технологий для анализа и обработки больших данных.
5. Использование программной среды MS Excel для анализа и обработки больших данных.
6. Использование программной среды Power BI для анализа и обработки больших данных.

6. Перечень основной и дополнительной учебной литературы

Основная литература:

1. Кравченко, Ю.А. Тенденции развития компьютерных технологий: учебное пособие / Ю.А. Кравченко, Э.В. Кулиев, Д.В. Заруба; Южный федеральный университет, Инженерно-технологическая академия. – Таганрог: Южный федеральный университет, 2017. – 107 с.: схем., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=493214> (дата обращения: 29.09.2020). – Библиогр. в кн. – ISBN 978-5-9275-2360-3. – Текст: электронный.

2. Лесковец, Ю. Анализ больших наборов данных / Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман; пер. с англ. А.А.Слинкина. - Москва: ДМК Пресс, 2016. - 498 с. - ISBN 978-5-97060-190-7. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1027845> (дата обращения: 06.10.2020). – Режим доступа: по подписке.

Дополнительная литература:

1. Анализ больших данных и приложение [Электронный ресурс] / Haitao Xiong .— : Aviation Industry Press, 2015 .— 326 с. — ISBN 978-7-5165-0956-2 .— Режим доступа: <https://lib.rucont.ru/efd/717979>

2. Форман, Д. Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - Москва: Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3. - Текст: электронный. - URL: <https://znanium.com/catalog/product/551044> (дата обращения: 06.10.2020). – Режим доступа: по подписке.

Рекомендуемая литература:

1. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013. – 464 с.

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://www.rusedu.info/> Сайт «Информационные технологии в образовании»
2. <http://nit.miem.edu.ru/> Сайт «Международная студенческая школа-семинар «Новые информационные технологии»».

8. Перечень информационных технологий

Перечень программного обеспечения: MSOffice (для создания отчетов), MS Excel, Power BI.

Информационные справочные системы: не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды МГОТУ

Рабочая программа и методическое обеспечение по дисциплине «Анализ больших данных».