

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

ИНСТИТУТ БИЗНЕСА И ДЕЛОВОГО АДМИНИСТРИРОВАНИЯ

Кафедра менеджмента

Утверждена
на заседании кафедры менеджмента
Протокол от «29» июня 2017 г. № 4

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1В.02 «Анализ больших данных»

по направлению подготовки 38.04.02 Менеджмент

направленность «Прорывные финансовые и цифровые технологии»

квалификация Магистр

очная форма обучения

Год набора - 2018

Москва, 2018 г.

Автор–составитель:

Кирилл Шилин, доктор технических наук, профессор кафедры эконометрии и мат.методов ЭМИТ РАНХиГС

Заведующий кафедрой менеджмента ИБДА, доктор экономических наук, профессор
Гапоненко Александр Лукич.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации
5. Методические указания для обучающихся по освоению дисциплины
6. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине
 - 6.1. Основная литература
 - 6.2. Нормативные правовые документы
 - 6.3. Интернет-ресурсы
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Цель и задачи освоения дисциплины

Целью освоения дисциплины «Анализ больших данных» является теоретическая и практическая подготовка студентов к работе с большими данными при решении экономических задач. Знания, полученные в результате освоения дисциплины, помогут при выборе методик анализа больших данных и визуализации полученных результатов. Все это необходимо выпускнику, освоившему программу магистратуры, для решения различных задач практической и научно-исследовательской деятельности.

Задачи освоения дисциплины состоят в формировании общепрофессиональных компетенций, позволяющих решать задачи профессиональной деятельности, связанные с анализом больших данных, с применением методов системного анализа и с использованием программы R-Studio.

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

1.1. Планируемыми результатами обучения по дисциплине «Анализ больших данных» являются знания, умения, владения, характеризующие уровни формирования компетенций и обеспечивающие достижение планируемых результатов освоения образовательной программы в целом. Перечень компетенций, формируемых в результате изучения дисциплины, приведен в таблице 1.

Таблица 1 – Формируемые компетенции

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-4	способность управлять организациями, подразделениями, группами (командами) сотрудников, проектными сетями	ПК-4.1	способен управлять организациями, подразделениями, группами (командами) сотрудников, проектными сетями
ПК ОС-13	способность принимать управленческие решения на основе анализа структурированных и неструктурированных данных	ПК ОС-13.1	способен принимать управленческие решения на основе анализа структурированных и неструктурированных данных
ОК-1	способность к абстрактному мышлению, анализу и синтезу	ОК-1.1	обладает абстрактным мышлением, анализом и синтезом

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

ОТФ/ТФ профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Способность принимать управленческие решения на основе анализа	ПК-4	На уровне знаний: - Основные понятия и терминологию в области технологий обработки и анализа больших

структурированных неструктурированных данных	и	<p>данных, понятие модели параллельной обработки данных MapReduce и ее практическую реализацию в Hadoop, основные API и типовые примеры программирования в Hadoop, стек Hadoop-технологий для распределенного хранения и обработки данных.</p> <p>- Языки программирования для обработки данных SAS Base и SAS Data Step 2, а также методы организации взаимодействия аналитической платформы SAS с распределенными хранилищами информации на основе Hadoop-стека технологий.</p> <p>- Семейство программных технологий SAS для обработки и анализа больших данных, включая программные продукты SAS для автоматической загрузки и предобработки больших данных, интерактивного исследования данных с использованием In-Memory подхода, построения прогнозных и описательных моделей, операционализации моделей, средств обработки разнородных сложно-структурированных данных, в том числе текстовых и сетевых (графовых).</p> <p>- Математические методы и модели представления данных для решения задач машинного обучения, статистического и интеллектуального анализа данных, прогнозирования временных рядов, задач оптимизации, используемых в программных решениях компании SAS, в том числе для распределенной работы со сложно-структурированными данными большого объема.</p> <p>Национ</p>
--	---	--

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

3 ЗЕ 108 ак. часов/81 астр. часа, количество академических часов, выделенных на контактную работу с преподавателем – 60/45 часа, на самостоятельную работу обучающихся – 42/31,5 часов, на контроль – 6/4,5 часов;

Место дисциплины в структуре ОП ВО

Освоение дисциплины Б1В.02 «Анализ больших данных» 1 курс, 1 семестр формирует у обучающихся компетенции, необходимые для подготовки магистра в соответствии с требованиями ФГОС ВО в области использования современных технологий для обработки, анализа и визуализации больших данных.

Входными требованиями к изучению дисциплины «Анализ больших данных» является наличие у студентов компетенций, сформированных на предыдущем уровне образования и непосредственно связанных с дисциплинами общепрофессионального цикла: «Теория вероятностей и математическая статистика», «Информатика и основы программирования».

Компетенции, приобретенные при освоении дисциплины «Анализ больших данных» могут быть использованы в практической и научно-исследовательской деятельности, а также

при выполнении дипломных работ студентов.

3. Структура и содержание дисциплины

3.1 Структура дисциплины

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.

Таблица 3 – Структура дисциплины

№ п/п	Наименование тем (разделов),	Объем дисциплины, час. /астр. часы						Форма текущего контроля успеваемости **, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР/ЭО,	ПЗ	КСР		
Очная форма обучения								
Тема 1	Основы анализа данных. Системы поддержки принятия решений	20	3		8	9		ДЗ
Тема 2	Проработка и очистка данных	20	3		9	8		ДЗ
Тема 3	Классификационный анализ без обучения. Кластерный анализ	20	4		8	8		ДЗ
Тема 4	Анализ взаимосвязей между переменными, ассоциативные правила	20	4		8	9		ДЗ
Тема 5	Классификационный анализ с обучением	20	4		9	8		ДЗ
Промежуточная аттестация								Зачет
Всего:		108/71	18/13,5	0	42/31,5	0	42/31,5	

3.2 Содержание дисциплины

3.2.1 Содержание теоретической части

Тема 1. Основы анализа данных. Системы поддержки принятия решений.

Сравнительный анализ OLAP и OLTP систем. Сферы их применения. Отличие информационного хранилища от баз данных. Принципы построения информационных хранилищ. Классификация информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Нормальная форма. Денормализация моделей данных. Правила Кодда. Зачем применяется денормализация моделей? Размерностные модели. Отличие таблицы фактов от размерностной таблицы. Характеристика стандартов DataMining.

Тема 2. Проработка и очистка данных.

Характеристика этапа ETL (Extracting Transforming and Loading). Задачи, решаемые DataMining. Предназначение и средства разведочный анализ данных. Характеристика диаграммы «ящик с усами». Операции, выполняемые при агрегировании данных. Примеры использования статистических пакетов для разведочного анализа. Скак выполняется сравнительный анализ графических средств анализа? Характеристика биржевых диаграмм. Использование диаграммы рассеяния.

Тема 3. Классификационный анализ без обучения. Кластерный анализ.

Основные характеристики кластера, «центроид» кластера, классификация методов кластерного анализа, примеры их применения в практической жизни. Меры близости, методы определения близости между кластерами, применение метода ближнего соседа, дальнего соседа, их сравнение. Характеристика метрик кластерного анализа. Содержание «дендограммы» и организацию ее применения. Использование статистических пакетов для решения задач кластерного анализа.

Тема 4. Анализ взаимосвязей между переменными, ассоциативные правила

Задачи использования ассоциативных правил, примеры задач использования ассоциативных правил, обобщенных правил. Понятие транзакции, примеры применения транзакций. Показатели, используемые для построения правила, алгоритмы построения ассоциативных правил, алгоритм apriori. Общая характеристика пакета Deductor. Использование пакета Deductor для решения задач интеллектуального анализа данных.

Тема 5. Классификационный анализ с обучением

Определение задачи классификации, методы решения задачи классификации, особенности решения задач классификации с обучением. Деревья классификации и их свойства, примеры алгоритмы построения деревьев. Алгоритм CART, примеры его использования. Что такое искусственный нейрон? Его состав, классификация решающих функций нейрона, их сравнительный анализ, искусственные нейронные сети. Что понимается под архитектурой нейронных сетей?

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Формы и методы текущего контроля успеваемости.

4.1.1. В ходе реализации дисциплины Б1В.02 «Анализ больших данных» используются следующие методы текущего контроля успеваемости обучающихся:

Тема (раздел)	Методы текущего контроля успеваемости
Тема 1	Домашнее задание 1
Тема 2	Домашнее задание 2
Тема 3	Домашнее задание 3
Тема 4	Домашнее задание 4
Тема 5	Домашнее задание 5

4.1.2. Зачет проводится с применением следующих методов (средств): в виде письменной контрольной работы/аналитического эссе (эсаймент)

4.2. Материалы текущего контроля успеваемости обучающихся

Варианты выполнения задания по Теме1:

- Дайте сравнительный анализ OLAP и OLTP систем. Сферы их применения
- Опишите - в чем отличие информационного хранилища от баз данных

Варианты выполнения задания по Теме 2:

- Каково предназначение и средства разведочный анализ данных? Дайте характеристику диаграммы «ящик с усами»
- Назовите какие операции выполняются при агрегировании данных. Приведите примеры использования статистических пакетов для разведочного анализа.

Выполненное задание оформляется:

- Процесс стратегического анализа на электронном носителе в Word.
- Презентация результатов работы в PowerPoint,

4.3. Оценочные средства для промежуточной аттестации.

Промежуточная аттестация проходит в форме защиты проекта.

В проекте студент формирует стратегию цифровой трансформации определенной организации. Анализируя существующие рынки, технологическую базу, действующие бизнес-модели и предлагая использования определенных технологий/бизнес-моделей, которые позволят перейти организации в цифровую эпоху

4.3.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы. Показатели и критерии оценивания компетенций с учетом этапа их формирования

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ОК-1	способность к абстрактному мышлению, анализу и синтезу	ОК-1.1	Знает основные понятия и основные методы теории анализа данных, интеллектуальной обработки данных; основные стандарты в области интеллектуального анализа данных.

4.3.2. Типовые оценочные средства.

При оценке зачета используется 100-балльная система. На экспертную оценку влияют следующие показатели:

1. Легкость для восприятия, представляемой им картины, описывающей будущее – 20%.
- 2. Реальность и его достижимость – 30%.
- 3. Понятные, четкие и убедительные, предлагаемые им стратегические управленческие решения – 50%.

Шкала оценивания

10- бальная шкала	Традиционн ая шкала	«Зачтено»/ «Не зачтено»	Определение
10	Отлично	Зачтено	Полные, глубокие и систематические знания анализа данных цифровой трансформации, полный и правильный ответ на теоретический вопрос, полное и правильное решение задачи.
9	Отлично	Зачтено	Глубокие и систематические знания анализа данных цифровой трансформации, правильный ответ на теоретический вопрос, правильное решение задачи.
8	Отлично	Зачтено	Систематические знания стратегического анализа данных цифровой трансформации, правильный ответ на теоретический вопрос, правильное решение задачи.
7	Хорошо	Зачтено	Систематические знания стратегического анализа данных цифровой трансформации, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи.
6	Хорошо	Зачтено	Систематические знания стратегического анализа данных цифровой трансформации, правильный ответ на теоретический вопрос с незначительными неточностями, правильное решение задачи с незначительными неточностями.
5	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации неполный, правильное решение задачи с незначительными неточностями.
4	Удовлетвори тельно	Зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации неполный, решение задачи содержит арифметические ошибки, не влияющие на правильность хода решения задачи.
3	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации неполный, решение задачи содержит идеологические ошибки.
2	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации неверный и/или решение задачи содержит идеологические ошибки.
1	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации неверный и решение задачи отсутствует.

10- бальная шкала	Традиционн ая шкала	«Зачтено»/ «Не зачтено»	Определение
0	Неудовлетво рительно	Не зачтено	Ответ на теоретический вопрос на тему стратегического анализа данных цифровой трансформации отсутствует и решение задачи отсутствует.

Предполагаемый перечень вопросов:

- Системы поддержки принятия решений. Хранилища данных.
2. Размерностные модели. OLAP-куб. Таблица размерностей. Таблица фактов.
 3. Сравнительный анализ OLAP и OLTP-систем.
 4. Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
 5. Этапы анализа данных. KDD.
 6. Data Mining. Средства обработки Data Mining
 7. Элементы математической статистики. Описательная статистика. Операции агрегирования данных.
 8. Графические средства анализа. Диаграмма рассеяния. Гистограмма.
 9. Начальные этапы KDD. ETL. Средства очистки и трансформации данных.
 10. Классификация методов предобработки и очистки данных.
 11. Методы борьбы с аномалиями. Ящечная диаграмма.
 12. Общая характеристика задач кластерного анализа.
 13. Метрики кластерного анализа.
 14. Методы определения близости между кластерами.
 15. Иерархическая кластеризация. Дендограмма.
 16. Метод k-средних.
 17. Ассоциативные правила. Свойства антимонотонности.
 18. Метрики построения ассоциативных правил.
 19. Алгоритм построения ассоциативных правил a'priori.
 20. Общая характеристика деревьев решений.
 21. Алгоритмы построения деревьев решений.
 22. Оценка качества классификации. Задачи классификации. ROC-кривая. Таблица сопряженности.
 23. Определение регрессионной модели. Логистическая регрессионная модель. Использование логистической модели для классификации.
 24. Нейронные сети. Перцептрон. Радиальные базисные сети.
 25. Использование карты Кохоннена для решения задач классификации.
 26. Общая характеристика QlikView.
 27. Общая характеристика DeductorAcademic.
 28. Общая характеристика средств ин

4.4. Методические материалы по проведению промежуточной аттестации

Зачет проводится в аудитории. Отсчет времени, отведенного на письменную работу, идет по завершении процедуры размещения студентов и раздачи заданий.

Студент обязан являться на письменный контроль в указанное в расписании время. В случае опоздания время, отведенное на письменный контроль знаний, не продлевается.

Мобильные телефоны должны быть выключены и убраны со столов, допускается использование калькуляторов, выполняющих только простые арифметические вычисления.

Во время проведения письменного контроля знаний студентам не разрешается

пользоваться учебными программами, справочниками и прочими источниками информации.

Использование материалов, а также попытка общения с другими студентами или иными лицами, в том числе с применением электронных средств связи, несанкционированные перемещения и т.п. являются основанием для удаления студента из аудитории и последующего проставления в ведомость оценки «неудовлетворительно».

Во время проведения письменного контроля знаний студентам разрешается покинуть аудиторию только при условии сдачи работы в объеме, выполненном к моменту выхода из аудитории. Дальнейшее продолжение работы запрещается.

5. Методические указания для обучающихся по освоению дисциплины

Все выступления, как преподавателя, так и студентов, могут сопровождаться наглядным материалом в виде презентаций, выполненных на компьютере, отрывками из документальных или художественных фильмов, иллюстрациями, прочим раздаточным материалом, что способствует более полному и глубокому освещению материала и легкости его усвоения слушателями. Развивается способность находить и выделять нужный информационный материал из разных видов источников, проводить его анализ и правильную подачу аудитории.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Основная литература.

1. Гмурман В.Е. Теория вероятностей и математическая статистика. Учебное пособие для бакалавров. - М.: Юрайт: ИД Юрайт, 2012. – 479 с.
2. Гусева Е.Н. Теория вероятностей и математическая статистика: учебное пособие. - Издательство: Флинта, 2011 г. – 220 с.- <http://rucont.ru/efd/246481>
3. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно-исследовательских работах. – М.: Форум, 2012. – 96 с.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник. -Издательство: ЮНИТИ-ДАНА, 2012. – 551 с.
5. Плохотникова К.Э. Основы эконометрики в пакете Statistica: учебное пособие для вузов. – М.: Вузовский учебник, 2011. – 297 с.
6. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / под ред. В.И. Ермакова. – М.: ИНФРА-М, 2012. – 287 с.
7. Шумак О.А., Гераськин А.В. Статистика: учебное пособие для студентов вузов. – М.: РИОР: ИНФРА-М, 2012. – 311 с.
8. 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.
9. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно-исследовательских работах. – М.: Форум, 2012. – 96 с.
10. Зарядов И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика. М.: Издательство Российского университета дружбы народов, 2010. - 207 с.
11. Зарядов И. С. Статистический пакет R: теория вероятностей и математическая статистика. М.: Издательство Российского университета дружбы народов, 2010. - 141 с.

12. Зорин А. В., Федоткин М. А. Введение в прикладной статистический анализ в пакете R. Учебно-методическое пособие. Нижний Новгород: Нижегородский гос. ун-т им. Н. И. Лобачевского, 2010. - 50 с.
13. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! - М.: ДМК Пресс, 2012. - 298 с

6.2. Нормативные правовые документы.

1. Конституция Российской Федерации. Принята на всенародном голосовании 12 декабря 1993 г. (с поправками от 30 декабря 2008 г.).
2. Гражданский кодекс Российской Федерации (часть вторая)» от 26 января 1996 г. № 14-ФЗ.
3. Европейская хартия местного самоуправления 1985 г. Ратифицирована Федеральным законом от 11 апреля 1998 года № 55-ФЗ «О ратификации Европейской хартии местного самоуправления» // Российская газета. 1998. 15 апр.
4. Концепция долгосрочного социально-экономического развития Российской Федерации на период до 2020 года, утверждена распоряжением Правительства Российской Федерации от 17 ноября 2008 г. № 1662-р.
5. Распоряжение Правительства РФ от 5 июля 2010 г. №1120-р «Об утверждении Стратегии социально-экономического развития Сибири до 2020 года».
6. Указ Президента Российской Федерации от 12 мая 2009 года №536 «Об Основах стратегического планирования в Российской Федерации».
7. Федеральный закон от 03.12.2011 № 392-ФЗ «О зонах территориального развития в Российской Федерации и о внесении изменений в отдельные законодательные акты Российской Федерации».
8. Федеральный закон от 16.10.2003 № 131-ФЗ «Об общих принципах организации местного самоуправления в Российской Федерации».
9. Федеральный закон от 21.07.2005 № 94-ФЗ «О размещении заказов на поставки товаров, выполнение работ, оказание услуг для государственных и муниципальных нужд».
10. Федеральный закон от 24.07.2007 № 209-ФЗ «О развитии малого и среднего предпринимательства в Российской Федерации».
11. Федеральный закон от 25.02.1999 № 39-ФЗ «Об инвестиционной деятельности в Российской Федерации, осуществляемой в форме капитальных вложений».
12. Федеральный закон от 25.12.2008 № 273-ФЗ «О противодействии коррупции».
13. Федеральный закон от 26.07.2006 № 135-ФЗ «О защите конкуренции».
14. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информатизации, информационных технологиях и защите информации».

6.5. Интернет-ресурсы.

<http://js.cdc.ranepa.ru/> - Центр цифровых компетенций РАНХиГС

<http://www.uecs.ru/> - Управление экономическими системами

www.nnir.ru / - Российская национальная библиотека

www.nns.ru / -Национальная электронная библиотека

www.rsi.ru / - Российская государственная библиотека

www.biznes-karta.ru / -Агентство деловой информации «Бизнес-карта»
[www. rbs.ru](http://www.rbs.ru) / - Информационное агентство «РосБизнесКонсалтинг»
www.aport.ru / - Поисковая система
www.rambler.ru / - Поисковая система
www.yandex.ru / - Поисковая система
www.busineslearning.ru / - Система дистанционного бизнес образования
www.test.specialist.ru / - Центр компьютерного обучения МГТУ им. Н. Э. Баумана
<http://www.consultant.ru/> - Консультант плюс
<http://www.garant.ru/> - Гарант

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Для обеспечения обучения студентов по дисциплине «Стратегический анализ» Академия располагает следующей материально-технической базой:

- лекционными аудиториями, оборудованными видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения, экраном и имеющие выход в сеть Интернет;
- помещениями для проведения семинарских и практических занятий, оборудованными учебной мебелью;
- библиотекой, имеющей рабочие места для обучающихся, оснащенные компьютерами с доступом к базам данных и сети Интернет;
- компьютерными классами.