МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
Обнинский институт атомной энергетики –
филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(ИАТЭ НИЯУ МИФИ)

ОТДЕЛЕНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ НАУК (О)

Одобрено на заседании
Ученого совета
ИАТЭ НИЯУ МИФИ
Протокол №25.8 от 25.08.2025

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ
для студентов
по освоению дисциплины

АНАЛИЗ БОЛЬШИХ ДАННЫХ

название дисциплины

для студентов направления подготовки

38.03.05 Бизнес-информатика

код и название направления подготовки

образовательная программа

Цифровая трансформация компании

Форма обучения: очная

г. Обнинск 2025 г.

ВВЕДЕНИЕ

Методические рекомендации для обучающихся по освоению дисциплины «Анализ больших данных» (рекомендуемый режим и характер учебной работы, в том числе в части выполнения самостоятельной работы) – комплекс рекомендаций и разъяснений, позволяющий обучающимся оптимальным образом организовать процесс изучения как теоретического учебного материала дисциплины, так и подготовки к практическим занятиям и/или лабораторным работам, в том числе проводимым с использованием активных и интерактивных технологий обучения.

Дисциплина «Анализ больших данных» является одной из профильных дисциплин в обеспечении профессионального становления будущего бизнес-информатика.

Цель дисциплины – формирование у обучающихся представлений и профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных. Данная цель соотносится с целью образовательной программы, в частности с технологией разработки специализированных программных систем, отвечающих за обработку больших данных.

Задачи изучения дисциплины:

- формирование у студентов представления о современных технологиях анализа больших данных;
- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- приобретение практических навыков работы с программой R-Studio.

Дисциплина реализуется в рамках части, формируемой участниками образовательных отношений, и относится к профессиональному модулю.

Дисциплина изучается на 3 курсе в 5 семестре.

Основными видами учебной работы по данной дисциплине являются лекции, практические занятия, самостоятельная работа обучающихся. Для успешного освоения дисциплины студенты необходимо изучить лекционный материал и рекомендуемую литературу, отработать изученный материал на практических занятиях, выполнить задания для самостоятельной работы.

1 Лекции

Лекция – это важный источник информации по каждой учебной дисциплине. Она ориентирует студента в основных проблемах изучаемого курса, направляет самостоятельную работу над ним.

Организация деятельности студента:

- По темам всех лекций имеются презентации.
- Отдельно старосте группы выдается список рекомендуемой литературы, имеющейся в библиотеке ИАТЭ и электронных библиотечных системах, для изучения тем по курсу.

Студент должен иметь лекционную тетрадь, где оформляет конспект лекций: кратко, схематично, последовательно фиксирует основные положения, выводы, формулировки, обобщения; помечает важные мысли, выделяет ключевые слова, термины. Проверка и уточнение терминов, понятий осуществляется студентом самостоятельно с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. При подготовке к занятиям необходимо обозначить вопросы, термины, материал, вызывающие трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удается разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации. Консультации проводятся как в очном, так и в дистанционном форматах.

Содержание лекционного курса по дисциплине «Анализ больших данных» представлено в таблице

Неделя	**Наименование раздела /темы дисциплины**	**Содержание**
1-4	I **Введение в большие данные**	
1-2	Основные понятия и терминология, классификация анализа больших данных	Основные вызовы больших данных (4V). Определение термина «большие данные». Представление о работе аналитика. Основные определения, термины, задачи анализа больших данных. Вопросы безопасности. 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.
3-4	Прикладные инструменты для работы с Big Data	Инструменты для обработки больших данных. Знакомство с языками и прикладными пакетами для обработки больших данных.
5-9	**II Статистические методы и программные средства для анализа больших данных**	

Неделя	Наименование раздела /темы дисциплины	Содержание
5-7	Аналитика больших данных. Методы анализа больших данных	Основные понятия математической статистики. данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализ, кластерный, дискриминантный, факторный анализ).
7-8	Информационное и программное обеспечение для работы по анализу больших наборов данных	Состав, виды и классификация программного обеспечения, используемого для анализа больших наборов данных. Функции, выполняемые программным обеспечением по анализу больших наборов данных. Программы, используемые в технологиях, используемых для обработки больших наборов данных.
8-9	Использование табличного процессора MS Excel для анализа больших наборов данных	Основные приемы работы в табличном редакторе: создание документа, ввод данных, редактирование, форматирование, больших наборов данных. Типы и форматы данных.
9-16	**III Использование баз данных для анализа больших наборов данных**	
9-10	Базы данных, общие сведения о РБД. Схема реляционной базы для больших наборов данных	Понятие и назначение баз данных. Система баз данных. Предметная область автоматизированных баз данных. Типы баз, используемые для работы с большими наборами данных: реляционные, иерархические, сетевые. Схема реляционной базы данных.
10-12	Система управления для работы с большими наборами данных	Понятие системы управления большими наборами данных. Виды СУБД, объекты, интерфейс. Понятие автоматизированной базы данных (АБД). Система баз больших наборов данных. Модели больших наборов данных: иерархическая, сетевая, реляционная. Виды систем управления большими наборами данных: реляционные, дедуктивные, экспертные, расширяемые, объектно-ориентированные, семантические, универсальные реляционные.
13-14	Программные инструменты для работы с большими наборами данных	Improvado - инструмент маркетинговой аналитики. Microsoft Power Query – для работы с импортируемыми большими наборами данных. Microsoft Power BI - инструмент бизнес-аналитики. Tableau Prep - известное программное обеспечение для подготовки данных. IBM SPSS Statistics для подготовки больших наборов данных. Qlik - инструмент

Неделя	Наименование раздела /темы дисциплины	Содержание
		бизнес-аналитики для БНД. DataWatch программное обеспечение для управления данными. Quest - средство для прогнозирования бизнес-данных. DataMeer - платформа на основе Saas для анализа больших наборов данных. Microstrategy - инструмент для анализа данных на рабочем столе, позволяет исследовать и анализировать данные. Paxata - самообслуживание и адаптивный инструмент для подготовки данных. Oracle - инструмент для самостоятельной подготовки данных.
15-16	OLAP и многомерные базы данных	OLAP - хранилище данных, предметно-ориентированное, привязанное ко времени и неизменяемое собрание данных для поддержки процесса принятия управляющих решений. OLAP - инструмент анализа БНД. OLAP - совокупность средств многомерного анализа данных. Определение OLAP - Online Analytical Processing, - оперативный анализ данных. OLAP - многомерное представление табличных данных. Операции с многомерным кубом -сечения, проекции, линейные таблицы. Иерархии и уровни OLAP. Архитектура OLAP-приложений. Технические проблемы многомерного хранения данных. OLAP-серверы.

Не стесняйтесь задавать вопросы преподавателю. Чем больше у Вас будет информации, тем свободнее и увереннее Вы будете себя чувствовать.

Базовые рекомендации:

- не старайтесь дословно конспектировать лекции, выделяйте основные положения, старайтесь понять логику лектора;
- точно записывайте определения, законы, понятия, формулы, теоремы и т.д.;
- передавайте излагаемый материал лектором своими словами;
- наиболее важные положения лекции выделяйте подчеркиванием;
- создайте свою систему сокращения слов;
- привыкайте просматривать, перечитывать перед новой лекцией предыдущую информацию;
- дополняйте материал лекции информацией;
- задавайте вопросы лектору;
- обязательно вовремя пополняйте возникшие пробелы.

Правила тактичного поведения и эффективного слушания на лекциях:

- слушать (и слышать) другого человека – это настоящее искусство, которое очень пригодится в будущей профессиональной деятельности;

- если преподаватель «скучный», но Вы чувствуете, что он действительно владеет материалом, то скука – это уже Ваша личная проблема (стоит вообще спросить себя, а настоящий ли Вы студент, если Вам не интересна лекция специалиста?).

Если Вы в чем-то не согласны (или не понимаете) с преподавателем, то совсем не обязательно тут же перебивать его и, тем более, высказывать свои представления, даже если они и кажутся Вам верными. Перебивание преподавателя на полуслове – это верный признак невоспитанности. А вопросы следует задавать либо после занятий (для этого их надо кратко записать, чтобы не забыть), либо, выбрав момент, когда преподаватель сделал хотя бы небольшую паузу, и обязательно извинившись.

2 Практические занятия (семинары)

Практические занятия являются важной частью учебного процесса в вузе. Они проводятся с целью закрепления лекционного материала, овладения понятийным аппаратом предмета, методами и приёмами исследования, изучаемыми в рамках учебной дисциплины. Главной целью такого рода занятий является научиться применению теоретических знаний на практике.

Содержание практических занятий по дисциплине «Анализ больших данных» представлено в таблице.

Неделя	**Наименование раздела /темы дисциплины**	**Содержание**
1-4	I **Введение в большие данные**	
1-2	Основные понятия и терминология, классификация анализа больших данных	Структура и задачи анализа больших наборов данных. Значение анализа больших наборов данных для инженерной и экономической работы. Формы представления информации по результатам анализа больших наборов данных. Понятие большого набора данных. Роль анализа больших наборов данных в развитии современных технологий управления в технической и экономической сферах деятельности. Функции управления с использованием анализа больших наборов данных. Место и роль анализа больших наборов данных в системах обработки данных, и системах управления. Информационные процессы, опосредованные применением методов анализа больших наборов данных. Классификация информационных технологий. Технология и методы обработки информации. Информационные ресурсы общества и предприятия. Виды и формы информационного обеспечения. Классификация информационных технологий по осуществлению анализа больших наборов данных. Системы обработки больших наборов данных

Неделя	Наименование раздела /темы дисциплины	Содержание
		и автоматизированные информационные системы. Методология и технология разработки информационных технологий анализа больших наборов данных. Принципы построения автоматизированных систем управления с применением анализа больших наборов данных
3-4	Прикладные инструменты для работы с Big Data	Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ. Сохранение данных в программе Excel. Преобразование и первичная обработка данных. Рассмотрение общей концепции и синтаксиса языка R (примеры). Базовое представление о Map Reduce и Hadoop.
5-9	**Статистические методы и программные средства для анализа больших данных**	
5-7	Аналитика больших данных. Методы анализа больших данных	Аналитика больших данных. Процесс аналитики. Стандарты жизненного цикла Big Data: CRISP-DM. Принципы и инструменты аналитики. Задачи и компетенции аналитиков Big Data. Big Data как рынок. Когнитивный анализ данных Введение в Data Mining – понятие, структура, составляющие и сопутствующие науки. • Задачи Data Mining и способы их решения. Классификация методов DM. • Области применения DM. • Классы систем DM. • Процесс накопления и анализа данных: Азбука когнитивного анализа. Аналитика больших данных. Математическая статистика Основные понятия статистики и дескриптивный анализ • Шкалы измерений. • Генеральная совокупность и выборка. • Нормальное распределение. Уровень статистической достоверности. • Свойства описательных статистик (Дескриптивный анализ) • Визуальное представление данных • Меры изменчивости Методы DATA MINING • Данные & знания

Неделя	Наименование раздела /темы дисциплины	Содержание
		• Типовые задачи Data Mining • Обучаемые и необучаемые задачи • Жизненный цикл проекта DM • Математический аппарат DM • Стандарты DM Методы анализа на графах • Случайные графы, безмасштабные графы, социальные сети – сети тесного мира. • Закономерности, методы кластеризации на графах. Прикладные инструменты анализа данных. Корреляция. • Готовые комплексные решения: Weka,RapidMiner, Knime, Orange IBM SPSS Modeler (в прошлом Clementine) • Инструменты визуализации: Tableau, • Фреймворки на JS, D3 • Корреляция. Понятие корреляции. Значимость коэффициента корреляции. Виды связи между переменными Многомерная регрессия и кластерный анализ
7-8	Информационное и программное обеспечение для работы по анализу больших наборов данных	Представление исходных данных в программе R-Studio (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах Excel и R-Studio: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы).
8-9	Использование табличного процессора MS Excel для анализа больших наборов данных	Организация расчетов в электронных таблицах с помощью пользовательских формул и функций, выполнение расчетов с использованием условий для обработки и анализа больших наборов данных. Графическое представление - построение диа-

Неделя	Наименование раздела /темы дисциплины	Содержание
		грамм, трех мерных графических систем для визуализации результатов анализа больших наборов данных в MS Excel.
9-16	**III Использование баз данных для анализа больших наборов данных**	
9-10	Базы данных, общие сведения о РБД. Схема реляционной базы для больших наборов данных	Типы баз, используемые для работы с большими наборами данных: реляционные, иерархические, сетевые. Схема реляционной базы данных.
10-12	Система управления для работы с большими наборами данных	Виды систем управления большими наборами данных: реляционные, дедуктивные, экспертные, расширяемые, объектно-ориентированные, семантические, универсальные реляционные.
13-14	Программные инструменты для работы с большими наборами данных	Возможности графического представления информации в программе R-Studio: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.
15-16	OLAP и многомерные базы данных	MOLAP (Multidimensional OLAP) - и детальные данные, и агрегаты хранятся в многомерной БД. ROLAP (Relational OLAP) - детальные данные остаются в реляционной БД. HOLAP (Hybrid OLAP) - детальные данные остаются в реляционной БД, а агрегаты хранятся в многомерной БД.

Семинарские занятия призваны научить студентов разбираться в проблемных вопросах анализа больших данных, ориентироваться в специальной литературе, самостоятельно работать с литературными и электронными источниками, научиться осуществлять поиск финансовой, маркетинговой и статистической информации, уяснять и уметь оценивать различные точки зрения. Также формируются навыки решения задач в области анализа различных предметных областей с использованием технологий больших данных.

Целью семинарских занятий для студентов, приступающих к изучению курса, является более глубокое знакомство с ключевыми теоретическими вопросами, изучаемыми на занятиях, а также овладение практическими приемами и методами решения задач в области анализа больших данных.

Основные задачи:

1) приобретение навыков научно-исследовательской работы на основе анализа текстов источников и применения различных методов исследования; 2) выработка умения самостоятельно и критически подходить к изучаемому материалу, включая библиографию и средства электронной информации (Интернет);

На практическом занятии обсуждаются теоретические положения изучаемого материала, уточняются позиции авторов научных концепций, ведется работа по осознанию студентами категориального аппарата изучаемой дисциплины, определяется и формулируется отношение учащихся к теоретическим проблемам науки, оформляется собственная позиция будущего специалиста. Форма работы – диалог: и студенты, и преподаватель вправе: задавать друг другу вопросы, которые возникли и могут возникнуть у них в процессе изучения и обсуждения материала, делиться своими сомнениями, наблюдениями, приводить доводы «за» и «против» той или иной позиции, обосновывать возможность применения на практике тех или иных теоретических положений.

Для подготовки к практическому занятию студентам рекомендуется:

- изучить вопросы, которые будут обсуждаться на занятии;
- изучить список основной и дополнительной литературы, где студенты могут найти ответы на вопросы, обратить внимание на категории, которыми оперирует автор, выписать основные понятия и систематизировать их;
- разработать блок-схему, в которой найдут отражение все изучаемые вопросы темы;
- составить развернутый план изучаемого материала, который может быть использован для ответа на занятии.

В начале занятия студенты под руководством преподавателя более глубоко осмысливают теоретические положения по теме занятия, раскрывают и объясняют основные положения публичного выступления. В процессе творческого обсуждения и дискуссии вырабатываются умения и навыки использовать приобретенные знания для различного рода ораторской деятельности.

Ввиду трудоемкости подготовки к практическому занятию преподаватель может предложить студентам алгоритм действий, рекомендовать еще раз внимательно прочитать записи лекций и уже готовый конспект по теме семинара, тщательно продумать свое устное выступление.

На практическом занятии каждый его участник должен быть готовым к выступлению по всем поставленным в плане вопросам, проявлять максимальную активность при их рассмотрении. Выступление должно строиться свободно, убедительно и аргументировано. Преподаватель следит, чтобы выступление не сводилось к репродуктивному уровню (простому воспроизведению текста), не допускается и простое чтение конспекта. Необходимо, чтобы выступающий проявлял собственное отношение к тому, о чем он говорит, высказывал свое личное мнение, понимание, обосновывал его и мог сделать правильные выводы из сказанного. При этом студент может обращаться к записям конспекта и лекций, непосредственно к первоисточникам, использовать знание художественной литературы и искусства, факты и наблюдения современной жизни и т. д. Вокруг такого выступления могут разгореться споры, дискуссии, к участию в которых должен стремиться каждый.

В заключение преподаватель подводит итоги практического занятия. Он может (выборочно) проверить конспекты студентов и, если потребуется, внести в них исправления и дополнения.

При изучении дисциплины используется значительное количество интерактивных методов обучения. Студенты привлекаются к активной со-творческой работе с преподавателем по поиску и подбору различных учебных материалов с использованием Интернет-ресурсов, а также формирования навыков организации профессионального взаимодействия с различными специалистами.

Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине, представлен в таблице.

№ п/п	Наименование темы дисциплины	Вид занятий (лекция, семинары, практические занятия)	Количество ак. час.	Наименование активных и интерактивных форм проведения занятий
I	**Введение в большие данные**			
1	Основные понятия и терминология, классификация анализа больших данных	лекция	1	Проблемная лекция с обсуждением содержания основных этапов развития электронной коммерции и ее технологий
2	Прикладные инструменты для работы с Big Data	лекция / практическое занятие	3	лекция-беседа, диспут, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ
II	**Статистические методы и программные средства для анализа больших данных**			
3	Аналитика больших данных. Методы анализа больших данных	лекция / практическое занятие	2	лекция-беседа, диспут, тесты, решение контрольных работ
4	Информационное и программное обеспечение для работы по анализу больших наборов данных	лекция / практическое занятие	3	Проблемная лекция и групповое обсуждение, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ
5	Использование табличного процессора MS Excel для анализа больших наборов данных	лекция / практическое занятие	2	Лекция-беседа и групповое обсуждение, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ
III	**Использование баз данных для анализа больших наборов данных**			
6	Базы данных, общие сведения о РБД. Схема реляционной базы для больших наборов данных	лекция / практическое занятие	2	Практико-ориентированная лекция-дискуссия, обсуждение содержания этапов планирования электронной

№ п/п	Наименование темы дисциплины	Вид занятий (лекция, семинары, практические занятия)	Количество ак. час.	Наименование активных и интерактивных форм проведения занятий
				коммерции и проектирования электронных организаций
7	Система управления для работы с большими наборами данных	лекция / практическое занятие	2	Лекция-беседа и групповое обсуждение, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ
8	Программные инструменты для работы с большими наборами данных	лекция / практическое занятие	3	Лекция-беседа и групповое обсуждение, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ
9	OLAP и многомерные базы данных	лекция / практическое занятие	3	Лекция-беседа, семинар с разбором конкретных ситуаций, тесты, решение контрольных работ

3 Самостоятельная работа обучающихся

Подготовка современного специалиста предполагает, что в стенах института он овладеет методологией самообразования, самовоспитания, самосовершенствования. Это определяет важность активизации его самостоятельной работы. С целью организации данного вида учебных занятий необходимо в первую очередь использовать материал лекций и семинаров. Лекционный материал создает проблемный фон с обозначением ориентиров, наполнение которых содержанием производится студентами на семинарских занятиях после работы с учебными пособиями, монографиями и периодическими изданиями.

В ходе изучения дисциплины студентам рекомендуется вечером того дня, когда было проведено занятие, прочитать лекцию или просмотреть решение задач на семинаре. За десять минут до начала лекции или семинара также прочитать предыдущую лекцию и просмотреть материалы семинара. Данные рекомендации обусловлены исследованием Эббингауза.

В соответствии с кривой забывания Эббингауза разработаны следующие режимы повторения для наилучшего запоминания:

Если есть два дня:

- первое повторение – сразу по окончании чтения;
- второе повторение – через 20 минут после первого повторения;
- третье повторение – через 8 часов после второго;

- четвёртое повторение – через 24 часа после третьего.

Если нужно помнить очень долго:

- первое повторение – сразу по окончании чтения;
- второе повторение – через 20-30 минут после первого повторения;
- третье повторение – через 1 день после второго;
- четвёртое повторение – через 2-3 недели после третьего;
- пятое повторение – через 2-3 месяца после четвёртого повторения

Самостоятельно изучается рекомендуемая литература, проводится работа с библиотечными фондами и электронными источниками информации, специальной литературой, статьями из профильных журналов. Реферируя и конспектируя наиболее важные вопросы, имеющие научно-практическую значимость, новизну, актуальность, делая выводы, заключения, высказывая практические замечания, выдвигая различные положения, студенты глубже понимают вопросы курса.

Подготовка к практическим занятиям, а также выполнение заданий для самостоятельной работы требует от студента навыков работы с литературными источниками:

- умение выделять главное в тексте;
- умение составлять опорную схему изученного материала, тезисный и развернутый план-конспект;
- свободное владение проработанным материалом;
- способность рассказать своими словами суть проблемы;
- умение объяснить и дать определение встречающимся в тексте новым научным терминам;
- умение находить в жизни ситуации, которые могут служить иллюстрацией теоретического материала, обсуждаемого на занятиях.

Своевременное и качественное выполнение самостоятельной работы базируется на соблюдении настоящих рекомендаций и изучении рекомендованной литературы. Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых работ и выпускной квалификационной работы.

Важной является готовность студента к восприятию в курсе сочетания философского, теоретического материала с конкретным практическим, направленным на освоение умений и навыков практической организации профессиональной деятельности в образовательном учреждении.

Подготовка к практическому занятию включает 2 этапа:

I - организационный;

II - закрепление и углубление теоретических знаний.

На первом этапе студент планирует свою самостоятельную работу, которая включает:

- уяснение задания на самостоятельную работу;
- подбор рекомендованной литературы;
- составление плана работы, в котором определяются основные пункты

предстоящей подготовки. Составление плана дисциплинирует и повышает организованность в работе.

Второй этап включает непосредственную подготовку студента к занятию. Начинать надо с изучения рекомендованной литературы. Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого материала, примеры, поясняющие его, а также разобраться в иллюстративном материале. Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам.

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику и тем самым проникнуть в творческую лабораторию автора.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

Важно развивать умение сопоставлять источники, продумывать изучаемый материал. Большое значение имеет совершенствование навыков конспектирования. Преподаватель может рекомендовать студентам следующие основные формы записи: план (простой и развернутый), выписки, тезисы. Результаты конспектирования могут быть представлены в различных формах.

План – это схема прочитанного материала, краткий (или подробный) перечень вопросов, отражающих структуру и последовательность материала. Подробно составленный план вполне заменяет конспект.

Конспект – это систематизированное, логичное изложение материала источника. Различаются четыре типа конспектов:

- план-конспект – это развернутый детализированный план, в котором достаточно подробные записи приводятся по тем пунктам плана, которые нуждаются в пояснении;

- текстуальный конспект – это воспроизведение наиболее важных положений и фактов источника;

- свободный конспект – это четко и кратко сформулированные (изложенные) основные положения в результате глубокого осмысливания материала. В нем могут присутствовать выписки, цитаты, тезисы; часть материала может быть представлена планом;

- тематический конспект – составляется на основе изучения ряда источников и дает более или менее исчерпывающий ответ по какой-то схеме (вопросу).

В процессе подготовки к занятиям рекомендуется взаимное обсуждение материала, во время которого закрепляются знания, а также приобретается практика в изложении и разъяснении полученных знаний, развивается речь.

При необходимости следует обращаться за консультацией к преподавателю. Идя на консультацию, необходимо хорошо продумать вопросы, которые требуют разъяснения.

Контрольные работы:

Подготовка к контрольным работам предполагает проработку лекционного материала, составление в рабочих тетрадях вспомогательных схем для наглядного структурирования материала с целью упрощения его запоминания. Необходимо обращать внимание на основную терминологию, классификации, отличительные особенности, наличие соответствующих связей между отдельными процессами в финансовой сфере. В зависимости от полноты освоения материала контрольная работа может проводиться в форме письменного ответа на теоретические вопросы и решения задач, а также в формате теста.

Подготовка доклада и презентации для участия в «круглом столе»

Основные этапы подготовки доклада

- выбор темы;
- консультация преподавателя;
- подготовка плана доклада;
- работа с источниками и литературой, сбор материала;
- написание текста доклада;
- оформление рукописи и предоставление ее преподавателю до начала доклада, что определяет готовность студента к выступлению;
- выступление с докладом, ответы на вопросы.

Тематика доклада предлагается преподавателем. Доклад должен быть подготовлен в печатной форме.

Технические требования к тексту доклада: шрифт 14, интервал 1,5, объем – 4-5 листов.

Текст доклада должен иметь титульный лист, оформленный в соответствии с образцом, имеющимся на кафедре, и содержать Ф.И.О. студента, Ф.И.О. преподавателя, название дисциплины, тему доклада, год выполнения, план доклада. Доклад должен содержать правильно оформленные ссылки на использованные источники и литературу. Доклад представляется в виде устного выступления и показа презентации.

Общие рекомендации к презентации:

- Шрифты с засечками читаются легче, чем гротески (шрифты без засечек);
- Для основного текста не рекомендуется использовать прописные буквы.

- Шрифтовой контраст можно создать посредством: размера шрифта, толщины шрифта, начертания, формы, направления и цвета.
- Цветовая гамма должна состоять не более чем из двух-трех цветов.
- Существуют не сочетаемые комбинации цветов.
- На полосе не должно быть больше семи значимых объектов, так как человек не в состоянии запомнить за один раз более семи пунктов чего-либо.
- Дизайн должен быть простым, а текст — коротким.

Рекомендации по оформлению и представлению на экране материалов различного вида.

- цвет шрифта и цвет фона должны контрастировать (текст должен хорошо читаться), но не резать глаза;
- тип шрифта: для основного текста гладкий шрифт без засечек (Arial, Tahoma, Verdana), для заголовка можно использовать декоративный шрифт, если он хорошо читаем;
- курсив, подчеркивание, жирный шрифт, прописные буквы рекомендуется использовать только для смыслового выделения фрагмента текста.
- рисунки, фотографии, диаграммы призваны дополнить текстовую информацию или передать ее в более наглядном виде;
- желательно избегать в презентации рисунков, не несущих смысловой нагрузки, если они не являются частью стилевого оформления;
- цвет графических изображений не должен резко контрастировать с общим стилевым оформлением слайда;
- иллюстрации рекомендуется сопровождать пояснительным текстом;
- если графическое изображение используется в качестве фона, то текст на этом фоне должен быть хорошо читаем;
- анимационные эффекты используются для привлечения внимания слушателей или для демонстрации динамики развития какого-либо процесса. В этих случаях использование анимации оправдано, но не стоит чрезмерно насыщать презентацию такими эффектами, иначе это вызовет негативную реакцию аудитории.
- стиль может включать: определенный шрифт (гарнитура и цвет), цвет фона или фоновый рисунок, декоративный элемент небольшого размера и др.;
- не рекомендуется использовать в стилевом оформлении презентации более 3 цветов и более 3 типов шрифта;
- оформление слайда не должно отвлекать внимание слушателей от его содержательной части;
- все слайды презентации должны быть выдержаны в одном стиле;
- информационных блоков не должно быть слишком много (3-6);
- рекомендуемый размер одного информационного блока — не более 1/2 размера слайда;

- желательно присутствие на странице блоков с разнотипной информацией (текст, графики, диаграммы, таблицы, рисунки), дополняющей друг друга;
- ключевые слова в информационном блоке необходимо выделить;
- информационные блоки лучше располагать горизонтально, связанные по смыслу блоки — слева направо;
- наиболее важную информацию следует поместить в центр слайда;
- логика предъявления информации на слайдах и в презентации должна соответствовать логике ее изложения.

Рекомендации к содержанию презентации.

- на слайдах презентации не пишется весь тот текст, который произносит докладчик (во-первых, в этом случае сам факт произнесения доклада теряет смысл, а во-вторых, длинный текст на слайде плохо воспринимается и только мешает слушанию и пониманию смысла);
- текст на слайде должен содержать только ключевые фразы (слова), которые докладчик развивает и комментирует устно;
- если презентация является основой устного доклада, то по европейским и американским правилам второй слайд должен содержать краткое перечисление всех основных вопросов, которые будут рассмотрены в докладе;
- все схемы и графики должны иметь названия, отражающие их содержание;
- в конце презентации представляется список использованных источников, оформленный по правилам библиографического описания.

Общие правила оформления презентации

Дизайн

Необходимо выбрать готовый дизайн или создать свой так, чтобы он соответствовал теме, не отвлекал слушателей.

Титульный лист

1. Название презентации.
2. Автор: ФИО, студента, место учебы, год.
3. Логотип (по желанию).

Второй слайд «Содержание» – список основных вопросов, рассматриваемых в содержании. Лучше оформить в виде гиперссылок (для интерактивности презентации).

Заголовки

1. Все заголовки выполнены в едином стиле (цвет, шрифт, размер, начертание).
2. В конце точка НИКОГДА не ставится (наверное, можно сделать исключение только для учеников начальной школы).
3. Анимация, как правило, не применяется.

Текст

1. Форматируется по ширине.

2. Размер и цвет шрифта подбираются так, чтобы было хорошо видно.
3. Подчеркивание НЕ используется, т.к. оно в документе указывает на гиперссылку.
4. Элементы списка отделяются точкой с запятой. В конце обязательно ставится точка.
5. На схемах текст лучше форматировать по центру.
6. В таблицах – по усмотрению автора.
7. Обычный текст пишется без использования маркеров списка:
8. Выделять главное в тексте другим цветом (желательно все в едином стиле).

Графика

1. Используются четкие изображения с хорошим качеством.
2. Лучше растровые изображения (в формате jpg) заранее обработать в любом графическом редакторе для уменьшения размера файла

Требования к оформлению диаграмм

1. У диаграммы должно быть название или таким названием может служить заголовок слайда;
2. Диаграмма должна занимать все место на слайде;
3. Линии и подписи должны быть хорошо видны.

Требования к оформлению таблиц

1. Название для таблицы;
2. Читаемость при невчитываемости;
3. Отличие шапки от основных данных.

Студент должен провести домашнюю репетицию устного выступления с докладом и удостовериться, что по времени доклад укладывается в отведенные для выступления 6-7 минут.
Домашняя (внеаудиторная) подготовка доклада оценивается до 1 балла, выступление и ответы на вопросы также до 1 балла Итого за выполнение данного задания студент может получить до 2-х баллов.

Индивидуальное домашнее задание (ИДЗ):

Выполнение индивидуального домашнего задания предполагает отработку навыков поиска, анализа и интерпретации материалов по тематике электронного бизнеса и электронной коммерции на основе использования нормативно-правовых актов. При выполнении данных заданий студент учится использовать в текущей деятельности нормативно-правовые акты для обоснования тех или иных решений в области электронного бизнеса.

В рамках самостоятельной подготовки к выполнению ИДЗ студент может изучить нормативно-правовую базу, изменения и дополнения к действующим нормативно-правовым актам, а также ознакомиться с необходимой ему информацией, размещенной на порталах регулирующих электронную коммерцию.

Формы организации самостоятельной работы обучающихся (темы, выносимые для самостоятельного изучения; вопросы для самоконтроля; типовые задания для самопроверки) представлены ниже.

На самостоятельное изучение выносятся следующие темы:

1. Процесс анализа больших данных.
2. Технологии анализа больших данных.
3. Научные проблемы в области больших данных.
4. Программы статистической обработки информации.
5. Жизненный цикл анализа больших данных, стандарты.
6. Когнитивный анализ данных.
7. Визуализация больших данных.
8. Технология MapRaduce. Hadoop.

Вопросы по указанным темам включены в контрольные работы и в перечень вопросов к зачету. Кроме того, в рамках выбранной темы студент может подготовить доклад для участия в круглом столе.

Вопросы для самоконтроля:

1. Большие данные (Big Data): современные подходы к обработке и хранению.
2. Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных.
3. Представления наборов данных.
4. Технологии KDD и Data Mining.
5. Подготовка данных к анализу.
6. Методика извлечения знаний. Data Mining.
7. Мультидисциплинарный характер Data Mining.
8. Причины распространения KDD и Data Mining.
9. Актуальность технологий Data Mining как средств обработки больших
10. объемов информации.
11. Программное обеспечение в области анализа данных.
12. Аналитические платформы: классификация и особенности применения.
13. Языки визуального моделирования.
14. Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила.
15. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил.
16. Частые предметные наборы и их обнаружение.
17. Алгоритм генерации ассоциативных правил.
18. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.
19. Определение кластеризации. Постановка задачи кластеризации.
20. Цели кластеризации в Data Mining.
21. Примеры кластеризации в различных областях.
22. Виды метрик. Шаги алгоритма. Меры расстояний.
23. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.

24. Применение классификации и регрессии. Обзор методов классификации и регрессии.
25. Статистические методы.
26. Методы, основанные на обучении, разнообразие подходов.
27. Постановка задач обучения по прецедентам.
28. Типы шкал: бинарные, номинальные, порядковые, количественные.
29. Типы задач: классификация, регрессия, прогнозирование, кластеризация. Примеры прикладных задач.
30. Постановка задачи кластеризации. Примеры прикладных задач.
31. Типы кластерных структур. Графовые алгоритмы кластеризации. Выделение связных компонент.
32. Кратчайший незамкнутый путь. Алгоритм ФОРЭЛ.
33. Функционалы качества кластеризации
34. Статистические алгоритмы: EM-алгоритм и Алгоритм k средних (k-means).
35. Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM
36. Самоорганизующаяся карта Кохонена. Применение для визуального анализа данных.
37. Искусство интерпретации карт Кохонена. Сети встречного распространения, их применение для кусочнопостоянной и гладкой аппроксимации функций
38. Агломеративная кластеризация, Алгоритм Ланса-Вильямса и его частные случаи.
39. Алгоритм построения дендрограммы. Определение числа кластеров. Свойства сжатия/растяжения, монотонности и редуктивности
40. Метод ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля.
41. Метод окна Парзена.
42. Метрические методы классификации в задаче восстановления регрессии. Обнаружение выбросов.
43. Понятия закономерности и информативности. Понятие логической закономерности. Эвристическое, статистическое, энтропийное определение информативности.
44. Асимптотическая эквивалентность статистического и энтропийного определения. Сравнение областей эвристических и статистических закономерностей.
45. Разновидности закономерностей: конъюнкции пороговых предикатов (гиперпараллелепипеды), синдромные правила, шары, гиперплоскости.
46. Градиентный алгоритм синтеза конъюнкций, частные случаи: жадный алгоритм, стохастический локальный поиск, стабилизация, редукция. Бинаризация признаков.
47. Решающие деревья для задач классификации и регрессии.
48. Линейный классификатор, непрерывные аппроксимации пороговой функции потерь. Связь с методом максимума правдоподобия.

49. Метод стохастического градиента и частные случаи: адаптивный линейный элемент ADALINE, персептрон Розенблатта, правило Хэбба.
50. Теорема Новикова о сходимости. Доказательство теоремы Новикова.
51. Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, "выбивание" из локальных минимумов.
52. Метод стохастического среднего градиента SAG.
53. Проблема мультиколлинеарности и переобучения, редукция весов (weight decay).
54. Байесовская регуляризация. Принцип максимума совместного правдоподобия данных и модели. Квадратичный (гауссовский) и лапласовский регуляризаторы.
55. Настройка порога решающего правила по критерию числа ошибок I и II рода.
56. Кривая ошибок (ROC curve). Алгоритм эффективного построения ROC-кривой. Градиентный метод максимизации AUC.
57.. Понятие опорных векторов. Рекомендации по выбору константы C. Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.
58. Способы конструктивного построения ядер. Примеры ядер. Обучение SVM методом активных ограничений.
59. SVM - регрессия. Метод релевантных векторов RVM. Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.
60. Метод наименьших квадратов, его вероятностный смысл и геометрический смысл.
61. Сингулярное разложение. Проблемы мультиколлинеарности и переобучения. Регуляризация.
62.. Гребневая регрессия. Лассо Тибширани, сравнение с гребневой регрессией.
63. Метод главных компонент и декоррелирующее преобразование Карунена - Лоэва, его связь с сингулярным разложением.
64. Линейные композиции, бустинг Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Взвешенное голосование.
65. Алгоритм AdaBoost. Процесс последовательного обучения базовых алгоритмов.
66. Теорема о сходимости бустинга. Базовые алгоритмы в бустинге. Решающие пни. Градиентный бустинг.
67. Стохастические методы: бэггинг и метод случайных подпространств. Случайные леса.
68. Оптимальный байесовский классификатор. Принцип максимума апостериорной вероятности. Функционал среднего риска.
69. Ошибки I и II рода. Теорема об оптимальности байесовского классификатора.
70. Оценивание плотности распределения: три основных подхода.
71. Наивный байесовский классификатор.

Примерные задания для самопроверки:

Задание 1. K-средних – это один из самых популярных методов кластеризации, позволяющий разделить исследуемое множество объектов на k кластеров (k – положительное число). Группировка в кластеры осуществляется на основе минимизации суммы корней расстояний между центроидами кластеров. Обобщённый алгоритм выглядит следующим образом:

1. Определить координаты центроидов.
2. Определить расстояние от каждого объекта до центроида.
3. Объединить объекты по признаку минимального расстояния.

Задание 2. Дан спектр S (простой числовой вектор) и вектор частот F, той же размерности что и S. Написать APL выражение для определения частоты, на которой спектр имеет максимальное значение.

Задание 3. Даны спектры S1 и S2 (простые числовые векторы) и вектор частот F, на которых оценивались спектры (той же размерности). Написать APL выражение для определения частот, на которых спектр S1 превышает спектр S2.

Задание 4. Дан числовой вектор P, содержащий значения цены товара, зафиксированные в различные моменты времени. Написать APL выражение, вычисляющее среднее по подмножеству значений, не превышающих заданный порог P.

4 Оценочные средства по дисциплине

Оценочные средства по дисциплине обеспечивают проверку освоения планируемых результатов обучения посредством мероприятий текущей и промежуточной аттестации.

4.1 Экзамен

а) типовые вопросы:

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
2. Проблемы анализа и обработки большого объема данных.
3. Базовые принципы обработки больших данных.
4. Определение модели. Свойства модели.
5. Аналитический подход к моделированию.
6. Информационный подход к моделированию.
7. Лица, участвующие в информационном моделировании.
8. Общая схема анализа.
9. Определение тиражирования знаний. Процесс построения модели.
10. Технологии обработки больших данных: NoSQL,
11. Технологии обработки больших данных: MapReduce,
12. Технологии обработки больших данных: Hadoop, R.
13. Методика извлечения знаний Knowledge Discovery in Databases (KDD). Этапы

KDD.
14. Data Mining. Постановка основных задач.
15. Машинное обучение.
15. Бизнес-решения с помощью алгоритмов Data Mining.
16. Классификация ПО в области Data Mining и KDD.
17. Типовая схема системы на базе аналитической платформы.
18. Понятие ассоциативного правила и транзакции.
19. Определение поддержки и достоверности.
20. Определение значимости и полезности ассоциативных правил, показатели их характеризующие.
21. Формальная постановка задачи кластеризации.
22. Цели кластеризации.
23. Основные шаги алгоритма k-means. Условие остановки алгоритма k-means. Преимущества и недостатки алгоритма k-means.
24. Кластеризация с помощью самоорганизующейся карты Кохонена
25. Этапы проведения классификации.
26. Обзор методов классификации и регрессии.
27. Задачи линейной и логистической регрессии.
28. Определение дерева решений. Структура дерева решений. Выбор атрибута разбиения в узле.

б) критерии оценивания компетенций (результатов):

Экзамен проводится устно и включает в себя ответ на три теоретических вопроса из различных разделов курса.

Освоение дисциплины оценивается по стобалльной системе, используемой в ИАТЭ НИЯУ МИФИ.

Допуск к экзамену по дисциплине осуществляется при количестве набранных в течение семестра баллов более 36.

За семестр студент может набрать от 36 до 60 баллов.

Минимальный балл за ответ на экзамене – 24, максимальный балл – 40.

Максимальная суммарная оценка за экзамен составляет 40 баллов с учетом того, что максимальная оценка работы в семестре по контрольным точкам составляет 60 баллов.

Ответ студента на экзамене оценивается по следующим критериям:

- правильность, полнота и логичность построения ответа;
- умение оперировать специальными терминами;
- использование в ответе дополнительного материала;
- умение иллюстрировать теоретические положения практическим материалом;
- владение актуальными статистическими данными по теме вопроса

Баллы	**Критерии оценки**
37-40	Студент должен:

	- продемонстрировать глубокое и прочное усвоение знаний программного материала; - исчерпывающе, последовательно, грамотно и логически стройно изложить теоретический материал; - правильно формулировать определения; - продемонстрировать умения самостоятельной работы с литературой; - уметь сделать выводы по излагаемому материалу
29-36	Студент должен: - продемонстрировать достаточно полное знание программного материала; - продемонстрировать знание основных теоретических понятий; - достаточно последовательно, грамотно и логически стройно излагать материал; - продемонстрировать умение ориентироваться в литературе; - уметь сделать достаточно обоснованные выводы по излагаемому материалу
25-28	Студент должен: - продемонстрировать общее знание изучаемого материала; - показать общее владение понятийным аппаратом дисциплины; уметь строить ответ в соответствии со структурой излагаемого вопроса; - знать основную рекомендуемую программой учебную литературу
менее 24	Студент демонстрирует: - незнание значительной части программного материала; - не владение понятийным аппаратом дисциплины; - существенные ошибки при изложении учебного материала; - неумение строить ответ в соответствии со структурой излагаемого вопроса; - неумение делать выводы по излагаемому материалу

4.2 Устный опрос

Оценочные средства представлены тематикой и вопросами, разработанными для обсуждения на семинарских занятиях.

Тема 1.

Вопросы:

1. Структура и задачи анализа больших наборов данных.
2. Значение анализа больших наборов данных для инженерной и экономической работы.
3. Формы представления информации по результатам анализа больших наборов данных. Понятие большого набора данных.

4. Роль анализа больших наборов данных в развитии современных технологий управления в технической и экономической сферах деятельности.
5. Функции управления с использованием анализа больших наборов данных

Тема 2.

Вопросы:

1. Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ.
2. Сохранение данных в программе Excel. Преобразование и первичная обработка данных.
3. Рассмотрение общей концепции и синтаксиса языка R (примеры).
4. Базовое представление о Map Reduce и Hadoop.

Тема 3.

Вопросы:

1. Аналитика больших данных. Процесс аналитики.
2. Стандарты жизненного цикла Big Data: CRISP-DM.
3. Принципы и инструменты аналитики.
4. Задачи и компетенции аналитиков Big Data.
5. Big Data как рынок.

б) критерии оценивания компетенций (результатов):

Устный опрос проходит в форме развернутой беседы – творческой дискуссии, основанной на подготовке всей группы по объявленной заранее теме при максимальном участии в обсуждении студентов группы. Как правило, один студент раскрывает один вопрос темы, давая наиболее полный ответ. Остальные делают дополнения, высказывают различные суждения и аргументацию, могут задавать вопросы друг другу и преподавателю. Преподаватель направляет ход дискуссии, обращая внимание на существующие научные и практические проблемы обсуждаемой темы, предлагая студентам найти собственное их решение.

в) описание шкалы оценивания:

Максимальная оценка за устное выступление и работу на семинарском занятии – 5 баллов.

5 баллов – студент дает полный ответ на поставленный вопрос, его речь свободна и грамотна, конспект не зачитывается, а используется лишь как опорный, студент делает важные дополнения по существу других вопросов, значительно проясняющие отдельные аспекты, которые не являются повторами, хорошо разбирается в обсуждаемом материале, демонстрирует знание источников, библиографии, различных точек зрения по изучаемой теме, умеет анализировать тексты, приходит к самостоятельным аргументированным выводам и отстаивает свою точку зрения, соблюдает нормы литературной речи.

2 балла – студент хорошо разбирается в обсуждаемом материале, демонстрирует умение критически анализировать источники и различные точки зрения по обсуждаемой проблеме, приходит к самостоятельным аргументированным выводам, однако не проявляет активность в работе группы на семинаре (готовится и отвечает только на один вопрос семинарского занятия).
1 балл – студент неполно владеет материалом, при изложении фактического материала допускает отдельные неточности, знает различные точки зрения по обсуждаемой проблеме, но возникают трудности с их анализом, умеет излагать собственную позицию, но не все выводы носят доказательный характер, при ответе активно пользуется конспектом вплоть до его зачитывания.

4.3 Рефераты (с последующим представлением докладов на «круглом столе»)

1. Понятие, сущность и ключевые признаки больших данных.
2. Роль и место больших данных в решении аналитических и исследовательских задач профессиональной деятельности.
3. Техника и технология Big Data
4. Техники больших данных. Консолидация данных. Визуализация.
5. Классификация. Кластеризация.
6. Регрессионный анализ.
7. Анализ ассоциативных правил.
8. Нейронные сети.
9. Технологии и инструменты больших данных.
10. Storm – система потоковой обработки.
11. Язык программирования R. Аналитика больших данных как корпоративный проект.
12. Аналитическая платформа Deductor Academic
13. Сущность и принцип работы аналитической платформы Deductor Academic. Основные функции и инструменты аналитической платформы Deductor Academic для целей анализа и исследования социально-экономических процессов и явлений в деятельности предприятий.
14. Моделирование социально-экономических процессов и явлений в деятельности предприятий с помощью платформы Deductor Academic.
15. Инструментарий прикладного компьютерного анализа и моделирования в Deductor Academic.
16. Метод факторного анализа как модель латентных переменных
17. Факторный анализ как метод понижения размерности пространства признаков
18. Экзогенные и эндогенные понятия в типологическом анализе
19. Технологии больших данных в финансах
20. Технологии больших данных в маркетинге
21. Технологии больших данных в логистике
22. Технологии больших данных в банковской сфере
23. Искусственный интеллект и его применение в управлении процессами

24. Машинное обучение.
25. Классификация клиентов организации с применением методов машинного обучения
26. Применение алгоритмов машинного обучения в задачах квантового анализа фондовой биржи
27. Применение машинного обучения для продвижения банковских продуктов
28. Применение методов машинного обучения при формировании оптимального портфеля ценных бумаг
29. Торговые роботы на основе обучения с подкреплением
30. Формирование моделей доходностей ценных бумаг на фондовой бирже на основе стратегий зависимости от новостей
31. Бизнес-аналитика на основе технологий больших данных

Показатели и критерии оценки реферата

Показатели оценки	**Критерии оценки**	**Баллы (max)**
1. Новизна реферированного текста	- актуальность проблемы и темы; - новизна и самостоятельность в постановке проблемы, в формулировании нового аспекта выбранной для анализа проблемы; - наличие авторской позиции, самостоятельность суждений	2
2. Степень раскрытия сущности проблемы	- соответствие плана теме реферата; - соответствие содержания теме и плану реферата; - полнота и глубина раскрытия основных понятий проблемы; - обоснованность способов и методов работы с материалом; - умение работать с литературой, систематизировать и структурировать материал; - умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы.	2
3. Обоснованность выбора источников	- круг, полнота использования литературных источников по проблеме; - привлечение новейших работ по проблеме (журнальные публикации, материалы сборников научных трудов и т.д.).	2
4. Соблюдение требований к оформлению	- правильное оформление ссылок на используемую литературу; - грамотность и культура изложения; - владение терминологией и понятийным аппаратом проблемы; - соблюдение требований к объему реферата;	2

	- культура оформления: выделение абзацев. - отсутствие орфографических и синтаксических ошибок, стилистических погрешностей; - отсутствие опечаток, сокращений слов, кроме общепринятых; - литературный стиль.	
5. Презентация на «круглом столе»	- качество и полнота презентации; - соблюдение требований к презентации по оформлению	2

Доклад – устное выступление студента, являющееся результатом его самостоятельной подготовки по заранее полученной теме и в соответствии с требованиями к самостоятельной работе студентов.

Выступление во время доклада, как правило, рассчитано на 6-7 минут, не может превышать установленное время, должно строго соответствовать объявленной теме. Приветствуются доклады с дополнительным использованием презентаций и мультимедийной техники.

Во время выступления студент может использовать свободную речь близко к тексту доклада, однако вправе зачитывать подготовленный им текст, демонстрируя владение материалом. Речь должна быть четкой, выразительной и эмоциональной. Обязательным элементом процедуры доклада является его обсуждение. Студентам группы предлагается задавать докладчику вопросы по теме доклада, вопросы может задавать и преподаватель. По окончании доклада организуется дискуссия.

Домашняя (внеаудиторная) подготовка доклада оценивается до 1 балла, выступление и ответы на вопросы на «круглом столе»- до 1 балла. Итого за выполнение данного задания студент может получить до 2 баллов.

Критерии оценки устного выступления.

2 балла (максимальная оценка) – выступление (доклад) отличается последовательностью, логикой изложения, легко воспринимается аудиторией, при ответе на вопросы выступающий демонстрирует глубину владения представленным материалом, ответы формулируются аргументированно, обосновывается собственная позиция в проблемных ситуациях.

1 балл – выступающий передает содержание проблемы, но не демонстрирует умение выделять главное, существенное, выступление воспринимается аудиторией достаточно сложно, ответы на вопросы поверхностные, либо вызывают у докладчика затруднение.

0 баллов – доклад краткий, поверхностный, несамостоятельный, докладчик не разбирается в сути вопроса, не может представить его в аудитории.

4.4 Тестовые задания

а) примеры тестов:

1. Какую часть мирового рынка Data Mining занимают услуги или консультации по эффективному внедрению этой технологии для решения актуальных бизнес-задач:
а) более 75%
б) около половины
в) менее 10% рынка

2. В основу программного продукта Cognos 4Thought положена технология:
а) множественной регрессии
б) нейронных сетей
в) деревьев решений

3. Достаточно высокая стоимость, невозможность добавлять свои функции, сложность подготовки данных, практическое отсутствие в интерфейсе терминов предметной области – это слабые стороны:
а) адаптация программного обеспечения под конкретную задачу
б) заказ готового решения у фирмы-разработчика
в) готового программного обеспечения

4. MAP-сплайны в системе STATISTICA – это:
а) непараметрическая процедура, в работе которой не используется никаких предположений об общем виде функциональных связей между зависимыми и независимыми переменными
б) процедура, опирающаяся на предположения о типе и накладывающая ограничения на класс зависимостей
в) параметрическая процедура, основанная на предположениях о виде функциональных связей между зависимыми и независимыми переменными

5. Готовые алгоритмы, полная конфиденциальность информации, техническая поддержка производителя, общение с другими пользователями пакета — это преимущества использования:
а) адаптация программного обеспечения под конкретную задачу
б) заказ готового решения у фирмы-разработчика
в) готового программного обеспечения

6. Cognos 4Thought предназначен для:
а) прогнозирования
б) моделирования
в) оба варианта верны
г) нет правильного ответа

7. Какое решение в большей мере требует наличия высококвалифицированных специалистов при внедрении и использования инструмента Data Mining:
а) адаптация программного обеспечения под конкретную задачу

б) заказ готового решения у фирмы-разработчика
в) использование готового программного обеспечения

8. Рабочее пространство STATISTICA Data Miner не включает такого элемента:
а) тестирование
б) анализ данных, моделирование
в) подготовка, преобразования и очистка данных

9. Вариант использования адаптированного программного обеспечения Data Mining:
а) имеет неоспоримые преимущества перед использованием готового программного обеспечения
б) всегда проигрывает перед использованием готового программного обеспечения
в) имеет как сильные, так и слабые стороны

10. Инструментальное средство для оперативного анализа данных и формирования отчетов по OLAP-технологии:
а) Cognos Impromptu
б) Cognos PowerPlay
в) Cognos Scenario

11. На этапе подготовки данных:
а) специалисты компании Разработчика подготавливают данные для их дальнейшего анализа
б) специалисты компании Разработчика и Заказчика подготавливают данные для их дальнейшего анализа
в) специалисты компании Заказчика подготавливают данные для их дальнейшего анализа

12. На каких этапах 4Thought поддерживает анализ данных:
а) сбор данных
б) преобразование данных
в) интерпретация модели
г) на всех этапах

13. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:
а) наличие терминов предметной области
б) сложность подготовки данных
в) полная конфиденциальность информации

14. Охарактеризуйте систему Cognos Scenario:
а) является средством оперативного анализа данных
б) формирует отчеты по OLAP–технологии
в) интеллектуальное инструментальное средство поиска данных

15. Преимуществом использования адаптированного программного обеспечения Data Mining по сравнению с готовыми программными продуктами и их самостоятельным использованием является:
а) адаптированность
б) не требуется дописывать программный код
в) сложность подготовки данных

16. Охарактеризуйте систему Cognos Scenario:
а) формирует отчеты по OLAP–технологии
б) позволяет руководителям выявлять скрытые тенденции и модели бизнеса
в) является средством оперативного анализа данных

17. На этапе первичного исследования данных:
а) всю работу осуществляет заказчик
б) со стороны заказчика может потребоваться максимальное участие
в) со стороны заказчика может потребоваться лишь минимальное участие

18. Модуль Oracle Data Mining доступен из таких редакций:
а) Personal Edition
б) Enterprise Edition
в) OneStandard Edition

19. Постановка бизнес-задачи – это этап, который:
а) формулирует конкретные бизнес-задачи, и они уже не могут быть изменены
б) формулирует конкретные бизнес-задачи, и они могут быть изменены в ходе прохождения именно этого цикла
в) формулирует конкретные бизнес-задачи, и они не могут быть изменены в ходе прохождения именно этого цикла

20. Архитектура хранилища типа «звезда» в Deductor называется:
а) сценарием
б) процессом
в) проектом

21. Что требуется от пользователя при работе с KXEN:
а) данные, которые необходимо проанализировать
б) выбор лучшей модели
в) тестирование модели

22. Oracle Data Mining поставляется как:
а) опция в Oracle Standard Edition
б) опция в Oracle Enterprise Edition
в) опция в любой редакции

23. Что требуется от пользователя при работе с KXEN:
а) выбор лучшей модели
б) тестирование модели
в) определение типа задачи, которую нужно решить

24. Последовательность действий, которые необходимо провести для анализа данных, называется в Deductor:
а) процессом
б) сценарием
в) этапом

25. Построение модели в KXEN можно охарактеризовать как:
а) итеративный процесс
б) набор таких шагов: подготовка модели, построение модели, тестирование модели
в) функцию предсказательного анализа в режиме on-line в формате «вопрос-ответ»

26. Один из компонентов, входящих в состав Deductor:
а) лаборатория BaseGroup Labs
б) многомерное хранилище данных Deductor Warehouse
в) лаборатория Base Labs

27. Охарактеризуйте квалификацию, которой требуется обладать пользователю для работы с KXEN:
а) пользователю не требуется обладать специальной квалификацией и знаниями в области анализа и статистики
б) пользователю требуется обладать специальной квалификацией и знаниями в области анализа, статистики и искусственного интеллекта
в) пользователю требуется обладать специальной квалификацией и знаниями в области анализа и статистики

28. Один из компонентов, входящих в состав Deductor:
а) лаборатория BaseGroup Labs
б) лаборатория Group Labs
в) аналитическое приложение Deductor Studio

29. В чем заключается основная особенность инструмента KXEN:
а) в практически полной автоматизации процесса построения моделей

б) в сложности построенных моделей
в) в возможности использования малого количества ретроспективных данных

30. Постановка задачи, построение оптимальной модели, понимание модели, применение результатов. Перечисленные выше этапы являются этапами:
а) традиционного процесса Data Mining
б) подхода KXEN
в) и того, и другого

Критерии оценивания компетенций (результатов):

Правильный ответ на каждое тестовое задание оценивается в 1 балл.

4.5. Контрольная работа

а) пример вопросов к контрольной работе
Контрольная работа «Введение в большие данные»
Вариант 1.

1. Структура и задачи анализа больших наборов данных.
2. Функции управления с использованием анализа больших наборов данных
3. Технология и методы обработки информации
4. Преобразование и первичная обработка данных
5. Принципы построения автоматизированных систем управления с применением анализа больших наборов данных

б) критерии оценивания компетенций (результатов):

Контрольные работы проводятся 1-2 раза в семестр на модульных неделях по расписанию, устанавливаемому учебно-методическим управлением. Они проводятся в форме типовых работ с вопросами, тестов или ином виде по выбору преподавателя с учетом объема изученного материала по курсу.

Контрольная работа может быть комбинированного типа: 3 вопроса, подразумевающих текстовые ответы и 5 тестовых вопросов с одним или несколькими вариантами ответа.

Оценивание студента проводится преподавателем независимо от наличия или отсутствия студента (по уважительной или неуважительной причине) на занятии. Студенту, пропустившему по уважительной причине контрольную модульную работу, предоставляется возможность отработки. Отработать занятие можно по согласованию с преподавателем в четко установленные сроки в соответствии с графиком консультаций преподавателя, который имеется в отделении социально-экономических наук.

Оценивается степень усвоения теоретических знаний по следующим критериям: правильность, полнота и логичность письменного ответа, способностью проиллюстрировать ответ примерами.

Максимальный балл за контрольную работу – 15. Каждый вопрос оценивается в 3 балла,

В случае использования вариантов заданий смешанного типа (вопросы, подразумевающие развернутый ответ, и тесты) баллы формируются следующим образом:

- 3 вопроса, максимальный балл за каждый вопрос – 4
- 2 тестовых задания, 1,5 балла за верно решенный тест, 0 баллов – если в тесте выбраны не все верные варианты, или среди выбранных вариантов ответа есть ошибочные

4.6. Индивидуальное домашнее задание

а) типовые задания (вопросы) – образец:

Выполнение индивидуального домашнего задания (ИДЗ) предполагает проведение студентами в группе из двух человек анализа ценных бумаг и производных финансовых инструментов с использованием технологий анализа больших данных и выбор оптимальных для компании или частного инвестора вариантов инвестирования с учетом максимально возможного для инвестора уровня рисков. Для проведения анализа студенты самостоятельно выбирают методы и модели. Рекомендуется использовать различные методы фундаментального и технического анализа рынка, а также методики формирования и оценки эффективности портфелей.

Вопросы к ИДЗ:

1. Проанализировать доходность акций (выбрать не менее 3 российских и 3 зарубежных компаний для анализа)
2. Проанализировать рынок российских и зарубежных корпоративных облигаций (выбрать не менее 3 российских и 3 зарубежных компаний для анализа)
3. Проанализировать доходность государственных ценных бумаг (облигации федерального займа, региональные и муниципальные облигации – не менее 2 субъектов РФ и муниципалитетов по данным за 5 лет с использованием технологий анализа больших данных)
4. Проанализировать рынок производных финансовых инструментов, выбрав не менее 3 инструментов для анализа, использовать дневные котировки за последние 5 лет (на основе использования методов анализа больших данных, в т.ч. – массивов с неполными, текстовыми данными и т.п.).
5. По каждому блоку анализа сделать выводы. Выбрать приемлемый для инвестора уровень риска и сформировать инвестиционный портфель с использованием технологий машинного обучения. Определить размер уплачиваемых инвестором налогов по операциям с ценными бумагами. Определить основные риски и возможную доходность данного портфеля.

б) критерии оценивания компетенций (результатов):

Выполнение индивидуального домашнего задания оценивается по следующим критериям:

1. правильность и полнота расчетов;
2. умение применять различные инструменты и модели анализа рынка ценных бумаг и производных финансовых инструментов для проведения анализа и обоснования выводов;
3. обоснованность выводов и предложений;
4. умение интерпретировать данные фундаментального и технического анализа более детального обоснования полученных выводов;
5. умение использовать специальную терминологию и выбирать методы анализа;
6. оригинальность работы (оценка оригинальности по системе «Антиплагиат. Вуз» - www.mephist. antiplagiat.ru должна быть не менее 70%)

в) описание шкалы оценивания:

Максимально за подготовку индивидуального домашнего задания студент может получить 10 баллов, максимальная оценка за каждый блок задания составляет 2 балла.

Оценка за выполненный блок формируется следующим образом:

2 балла - все расчеты выполнены верно, методики использованы корректно, выводы обоснованы, для обоснования выводов и предложений представлен детальный анализ дополнительных форм внутренней отчетности, терминология используется корректно и ее использование полностью соответствует содержанию блока, полученные рекомендации обоснованы анализом действующего законодательства.

1,5 балла - все расчеты выполнены верно, выводы обоснованы, для обоснования выводов и предложений представлен детальный анализ дополнительных форм отчетности, терминология используется корректно и ее использование полностью соответствует содержанию блока, полученные рекомендации обоснованы анализом действующего законодательства, но имеются незначительные неточности в выводах, или рассмотрены не все аспекты проблемы.

1 балл - расчеты выполнены в целом верно, выводы обоснованы, терминология используется корректно и ее использование полностью соответствует содержанию блока, однако для обоснования рекомендаций не достаточно полно используется дополнительная отчетность и анализ действующего законодательства.

0,5 балла - в расчетах присутствуют ошибки, выводы носят описательный характер. Рекомендации схематичны, не представлен анализ дополнительных форм отчетности и действующего законодательства для

обоснования рекомендаций.

0 баллов – блок задания не выполнен, или выполнен с грубыми ошибками: ошибки в расчетах и использовании формул, отсутствие выводов или их несоответствие практике.

5 Итоговая аттестация по дисциплине

Итоговая аттестация по дисциплине является интегральным показателем качества теоретических и практических знаний и навыков обучающихся по дисциплине и складывается из оценок, полученных в ходе текущей и промежуточной аттестации.

Текущая аттестация в семестре проводится с целью обеспечения своевременной обратной связи, для коррекции обучения, активизации самостоятельной работы обучающихся.

Промежуточная аттестация предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины.

Текущая аттестация осуществляется два раза в семестр:

-контрольная точка № 1 (КТ № 1) – выставляется в электронную ведомость не позднее 8 недели учебного семестра. Включает в себя оценку мероприятий текущего контроля аудиторной и самостоятельной работы обучающегося по разделам/темам учебной дисциплины с 1 по 8 неделю учебного семестра.

- контрольная точка № 2 (КТ № 2) – выставляется в электронную ведомость не позднее 16 недели учебного семестра. Включает в себя оценку мероприятий текущего контроля аудиторной и самостоятельной работы обучающегося по разделам/темам учебной дисциплины с 9 по 16 неделю учебного семестра.

Результаты текущей и промежуточной аттестации подводятся по шкале балльно-рейтинговой системы.

Этап рейтинговой системы / Оценочное средство	**Неделя**	**Балл**	
		Минимум*	Максимум
Текущая аттестация	**1-16**	**36**	**60**
Контрольная точка № 1	**7-8**	**18**	**30**
Контрольная работа	8	9	15
Опрос	1-8	3	5
Индивидуальное домашнее задание	7	6	10
Контрольная точка № 2	**15-16**	**18**	**30**
Реферат +доклад на «круглом столе»	16	6	10
Тестовые задания	9-15	9	15
Опрос	9-16	3	5
Промежуточная аттестация	**-**	**24**	**40**

Экзамен		24	40
Итого по дисциплине		**60**	**100**

* - Минимальное количество баллов за оценочное средство – это количество баллов, набранное обучающимся, при котором оценочное средство засчитывается, в противном случае обучающийся должен ликвидировать появившуюся академическую задолженность по текущей или промежуточной аттестации. Минимальное количество баллов за текущую аттестацию, в т. ч. отдельное оценочное средство в ее составе, и промежуточную аттестацию составляет 60% от соответствующих максимальных баллов.

Процедура оценивания знаний, умений, владений по дисциплине включает учет успешности по всем видам заявленных оценочных средств.

Устный опрос проводится на каждом практическом занятии и затрагивает как тематику прошедшего занятия, так и лекционный материал. Применяется групповое оценивание ответа или оценивание преподавателем.

Темы рефератов распределяются на первом занятии, готовые рефераты докладываются на занятиях в сопровождении презентаций в соответствии с установленным преподавателем графиком.

По окончании освоения дисциплины проводится промежуточная аттестация в виде экзамена, что позволяет оценить совокупность приобретенных в процессе обучения компетенций. При выставлении итоговой оценки применяется балльно-рейтинговая система оценки результатов обучения.

Экзамен предназначен для оценки работы обучающегося в течение всего срока изучения дисциплины и призван выявить уровень, прочность и систематичность полученных обучающимся теоретических знаний и умений, способности приводить примеры практического использования знаний (например, применять их в решении практических задач), приобретения навыков самостоятельной работы, развития творческого мышления.

Оценка сформированности компетенций на экзамене для тех обучающихся, которые пропускали занятия и не участвовали в проверке компетенций во время изучения дисциплины, проводится после индивидуального собеседования с преподавателем по пропущенным или не усвоенным обучающимся темам с последующей оценкой самостоятельно усвоенных знаний на экзамене.

Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет сумму баллов, заработанных обучающимся при выполнении заданий в рамках текущей и промежуточной аттестации

Сумма баллов	**Оценка по 4-х балльной шкале**	**Оценка ECTS**	**Требования к уровню освоения учебной дисциплины**
90-100	5- «отлично»/ «зачтено»	A	Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил про-

			граммный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы
85-89	4 - «хорошо»/ «зачтено»	B	Оценка «хорошо» выставляется обучающемуся, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос
75-84		C	
70--74		D	
65-69	3 - «удовлетворительно» / «зачтено»		Оценка «удовлетворительно» выставляется обучающемуся, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала
60-64		E	
0-59	2 - «неудовлетворительно»/ «не зачтено»	F	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится обучающимся, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине