Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Макаренко Елена Николаевна
Должность: Ректор
Дата подписания: 30.08.2026 10:49:46
Уникальный программный ключ:
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Начальник
учебно-методического управления
Т.К. Платонова
«25» мая 2026 г.

Рабочая программа дисциплины
Извлечение и анализ интернет-данных

Направление подготовки
01.03.05 Статистика

Направленность (профиль) программы бакалавриата
01.03.05.01 Анализ больших данных

Для набора 2026 года

Квалификация
Бакалавр

КАФЕДРА **Статистики, эконометрики и оценки рисков**

Распределение часов дисциплины по семестрам / курсам

Семестр (<Курс>.<Семестр на курсе>)	**6 (3.2)**		**7 (4.1)**		Итого	
Недель	16 2/6		16 1/6			
Вид занятий	УП	РП	УП	РП	УП	РП
Лекции	32	32	16	16	48	48
Лабораторные	32	32	32	32	64	64
Итого ауд.	64	64	48	48	112	112
Контактная работа	64	64	48	48	112	112
Сам. работа	80	80	60	60	140	140
Часы на контроль			36	36	36	36
Итого	144	144	144	144	288	288

ОСНОВАНИЕ

Учебный план утвержден учёным советом Университета (протокол № 9 от 03.03.2026 г.).

Программу составил(и): к.э.н., доцент, Бондаренко Г.А.

Зав. кафедрой: д.э.н., профессор Л.И. Ниворожкина

Методический совет: к.э.н., доцент О.В. Андреева

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Цель освоения дисциплины - формирование у обучающихся практических навыков поиска, извлечения и анализа данных из различных источников сети Интернет, необходимую для решения научных и профессиональных задач (в том числе на основе системного подхода).

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПК-6. Способен осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных

ПК-4. Способен формировать источники больших данных

В результате освоения дисциплины обучающийся должен:

Знать:

- основные источники больших данных сети Интернет (соотнесено с индикатором ПК-4.1);
- основные этапы проведения статистического наблюдения, как основы получения информации для анализа больших данных, извлеченных из сети Интернет (соотнесено с индикатором ПК-6.1).

Уметь:

- подбирать, анализировать и обобщать большие данные из Интернет источников (соотнесено с индикатором ПК-4.2);
- осуществлять анализ результатов проведенного аналитического исследования с использованием технологий больших данных (соотнесено с индикатором ПК-6.2)

Владеть:

- навыками формирования источников больших данных, извлеченных из сети Интернет (соотнесено с индикатором ПК-4.3);
- навыками подготовки статистического инструментария в целях получения первичной информации из сети Интернет для осуществления обработки и анализа данных (соотнесено с индикатором ПК-6.3).

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Теоретические и методологические основы анализа данных, извлеченных из сети Интернет

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
1.1	Базовые понятия и основные задачи анализа данных, извлеченных из сети Интернет. Этапы и основные методы анализа данных, извлеченных из Интернет источников. Эволюция методов обработки и анализа данных из Интернет ресурсов.	Лекционные занятия	6	4	ПК-6 ПК-4
1.2	Алгоритмы предварительной подготовки данных, извлеченных из сети Интернет. Понятие "данные". Набор данных и их атрибутов. Типы наборов данных. Измерения и шкалы. Методы и способы получения первичной информации из сети Интернет. Шумы и выбросы в данных. Пропуски данных и методы их восстановления. Задачи отбора признаков. Несбалансированные выборки. Базы и витрины данных.	Лекционные занятия	6	4	ПК-6 ПК-4
1.3	Концептуальные подходы к поиску информации в сети Интернет. Классический поиск. Нечёткий поиск. Web Mining. Анализ изображений, аудио и видео. Интеллектуальный веб-поиск. Классификация подходов к поиску информации	Лекционные занятия	6	6	ПК-6 ПК-4
1.4	Парсинг открытых данных в различных форматах. Понятие парсинга открытых данных. Нормативно-правовая база парсинга открытых данных. Основные инструменты web scraping. Области применения инструментов парсинга: сбор данных для исследования социально-экономических явлений и процессов, извлечение контактной информации; поиск работы или сотрудников, мониторинг цен.	Лекционные занятия	6	4	ПК-6 ПК-4
1.5	Методы поиска информации в сети Интернет Булев поиск. Векторно-пространственный, вероятностный и языковой поиск. Хеширование и деревья поиска. Методы Web Content Mining. Машинное обучение для поиска по запросу. Латентно-семантическое индексирование. Распознавание образов для поиска изображений. Методы поиска аудио и видео. Методы интеллектуального веб-поиска. Оценка достоверности информации. Классификация методов поиска информации	Лекционные занятия	6	4	ПК-6 ПК-4
1.6	Алгоритмы поиска информации в сети Интернет Построение и сжатие индекса. Ранжирование документов по степени соответствия запросу. Алгоритм PageRank. Алгоритмы классификации. Алгоритмы кластеризации. Автоматическое реферирование. Поиск семантических связей в тексте. Оценка	Лекционные занятия	6	4	ПК-6 ПК-4

	смыслового подобия. Поиск ответов на вопрос. Классификация алгоритмов поиска информации				
1.7	Лабораторная работа 1. Обзор основных открытых баз и витрин данных в целях осуществления анализа данных Базы данных Федеральной службы государственной статистики (Росстат), органов исполнительной власти на разных уровнях территориальной агрегации, Единой межведомственной информационно-статистической системы, Всероссийского центра исследования общественного мнения (ВЦИОМ), Центрального банка РФ, Национального агентства финансовых исследований, Фонда общественного мнения, портала открытых данных РФ, kaggle.com и т.п. Подготовка отчета по лабораторной работе 1.	Лабораторные занятия	6	6	ПК-6 ПК-4
1.8	Лабораторная работа 2. Извлечение и анализ данных из открытых источников. Подготовка отчета по лабораторной работе 2.	Лабораторные занятия	6	6	ПК-6 ПК-4
1.9	Лабораторная работа 3. Классификация и кластеризация данных с помощью различных инструментальных средств (LibreOffice, Gretl) Подготовка отчета по лабораторной работе 3.	Лабораторные занятия	6	6	ПК-6 ПК-4
1.10	Лабораторная работа 4. Методы графического представления и визуализации данных. Подготовка отчета по лабораторной работе 4.	Лабораторные занятия	6	4	ПК-6 ПК-4
1.11	Информационное пространство как среда анализа данных. Понятие и структура информационного пространства. Элементы структуры информационного пространства. Понятие "показатель". Пространственная интерпретация понятия "показатель". Содержание экономических показателей. Построение системы показателей.	Самостоятельная работа	6	10	ПК-6 ПК-4
1.12	Концепции организации хранения данных. Концепция информационного хранилища. Концепция централизованного хранилища данных. Концепция распределенного хранилища данных. Концепция автономных витрин данных. Концепция единого интегрированного хранилища и многих витрин данных.	Самостоятельная работа	6	14	ПК-6 ПК-4
1.13	База метаданных информационного хранилища. основные понятия о базе метаданных. Модели метаданных по измерениям. Содержание компонентов размерностной модели метаданных информационного хранилища.	Самостоятельная работа	6	16	ПК-6 ПК-4
1.14	Модели данных информационного хранилища. Понятие, структура и схемы представления данных модели информационного хранилища	Самостоятельная работа	6	20	ПК-6 ПК-4
1.15	Роль и место интеллектуального анализа данных сети Интернет в процессе принятия управленческих решений. Внутренние и внешние информационные ресурсы. Границы информационного пространства лица, принимающего решение (ЛПР). Информационно-аналитические системы. Аналитическая подготовка принятия решений. Проблема извлечения данных из разнородных источников. Организация хранения полученной информации.	Самостоятельная работа	6	20	ПК-6 ПК-4

Раздел 2. Инструментальные средства интеллектуального анализа данных, извлеченных из сети Интернет

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
2.1	Технологии извлечения, преобразования и загрузки данных. Понятие о гибкой архитектуре данных. Основные свойства открытых систем. Элементы электронного обмена данными. Повышение качества данных и приведение их к одной структуре. Оценка качества данных.	Лекционные занятия	6	4	ПК-6 ПК-4
2.2	Универсальные поисковые системы в сети Интернет. История развития универсальных поисковых систем. Характеристики и возможности универсальных поисковых систем. Сравнительный анализ поисковых систем. Перспективные направления развития универсальных поисковых систем. Схема выполнения задач поиска информации универсальными поисковыми системами	Лекционные занятия	6	2	ПК-6 ПК-4
2.3	Лабораторная работа 5. Реализация и интерпретация результатов многомерного статистического анализа данных, извлеченных из Интернет источников с помощью различных инструментальных средств (LibreOffice, Gretl) Подготовка отчета по лабораторной работе 5.	Лабораторные занятия	6	10	ПК-6 ПК-4
2.4	Подготовка к промежуточной аттестации	Зачет	6	0	ПК-6 ПК-4

Раздел 3. IT-анализ данных, извлеченных из сети Интернет

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
3.1	Аналитические платформы Data Mining, используемые для анализа	Лекционные	7	4	ПК-6

	данных, извлеченных из Интернет источников: основные возможности, сфера применения, этапы решения задач. Импорт/ экспорт данных. Визуализация результатов. Возможности автоматизации отдельных этапов анализа данных, извлеченных из Интернет источников.	занятия			ПК-4
3.2	Лабораторная работа 1. Реализация и интерпретация процедуры анализа временных рядов и прогнозирования с помощью различных инструментальных средств (LibreOffice, Gretl) Подготовка отчета по лабораторной работе 1.	Лабораторные занятия	7	8	ПК-6 ПК-4
3.3	Подходы к выполнению анализа данных Интернет ресурсов средствами информационных технологий (IT-анализа). Классификация IT-анализа о режиму и темпу. Задачи и содержание оперативного (OLAP) анализа. Типы многомерных OLAP систем.	Самостоятельная работа	7	10	ПК-6 ПК-4
3.4	Типы инструментальных средств создания и поддержки анализа данных Интернет ресурсов. Назначение инструментальных средств, их классификация. Аналитические инструментальные средства пакетов прикладных программ широкого применения LibreOffice. Специализированные средства и их характеристика. Блоки интегрированных экономических информационных систем.	Самостоятельная работа	7	10	ПК-6 ПК-4

Раздел 4. Практика применения интеллектуальных технологий анализа данных, извлеченных из сети Интернет

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
4.1	Применение данных, извлеченных из сети Интернет для решения бизнес-задач. Отбор данных для анализа. Предварительная подготовка данных. Понижение размерности, сглаживание аномалий, фильтрация, группировка. Анализ обеспечения ресурсами. Анализ возможных прерываний бизнес-процессов. Определение верхних границ цен. Анализ стратегической позиции предприятия	Лекционные занятия	7	4	ПК-6 ПК-4
4.2	Применение данных сети Интернет в маркетинговой деятельности и в сфере логистики. Анализ разрыва. Портфолио-анализ. Анализ маржинальной прибыли. Сравнительные расчеты. Анализ цепочек и издержек логистических процессов.	Лекционные занятия	7	4	ПК-6 ПК-4
4.3	Финансовый и инвестиционный данных сети Интернет. Анализ потоков платежей. Финансовая "паутина". Функционально-стоимостной анализ. Анализ ставки дисконтирования с поправкой на риск.	Лекционные занятия	7	4	ПК-6 ПК-4
4.4	Лабораторная работа 2. Реализация и интерпретация результатов анализа данных сети Интернет для решения основных бизнес задач. Подготовка отчета по лабораторной работе 2.	Лабораторные занятия	7	8	ПК-6 ПК-4
4.5	Лабораторная работа 3. Реализация и интерпретация результатов анализа данных сети Интернет в маркетинговой деятельности и в сфере логистики. Подготовка отчета по лабораторной работе 3.	Лабораторные занятия	7	8	ПК-6 ПК-4
4.6	Лабораторная работа 4. Реализация и интерпретация результатов финансового и инвестиционного анализа данных сети Интернет. Подготовка отчета по лабораторной работе 4	Лабораторные занятия	7	8	ПК-6 ПК-4
4.7	Содержание и методы экономического анализа и прогнозирования бизнес-процессов (деятельности предприятий). Содержание экономического анализа. Классификация методов анализа. Аннотация содержания методов анализа в экономической предметной области.	Самостоятельная работа	7	10	ПК-6 ПК-4
4.8	Функционально-стоимостной анализ инвестиционной и инновационной деятельности. Субъективные оценки проектов путем составления иерархии целей, их взвешивания, составления таблиц функций и определения альтер-натив реализации функций. Расчет полезности и формирование последовательности в матрице ценности целей. Анализ чувствительности полученных данных к изменению весовых коэффициентов целей, оценкой и выдачей результата.	Самостоятельная работа	7	10	ПК-6 ПК-4
4.9	Информационный обмен на предприятии, связанный с аналитической работой. Сущность аналитической работы на предприятии. Место интеллектуального анализа в цепочке принятия управленческих решений. Источники информации на предприятии для проведения ИАД. Потребители аналитической информации. Форма представления аналитической информации. Документооборот на предприятии.	Самостоятельная работа	7	10	ПК-6 ПК-4
4.10	Подходы, используемые при автоматизации процессов экономического анализа. Рассмотрение и обоснование необходимость применения информационных технологий при выполнении аналитических работ с имеющимися в распоряжении ЛПР данными для выработки	Самостоятельная работа	7	10	ПК-6 ПК-4

	адекватных сложившимся ситуациям решений.				
4.11	Подготовка к промежуточной аттестации	Экзамен	7	36	ПК-6 ПК-4

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущего контроля и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1	Марков А. С., Лисовский К. Ю.	Базы данных. Введение в теорию и методологию: учеб.	М.: Финансы и статистика, 2006	Библиотека РГЭУ (РИНХ) / 50 экз.
2		Журнал "Вопросы статистики"		Читальный зал РГЭУ (РИНХ)
3	Чубукова И. А.	Data Mining: учебное пособие	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)| Бином. Лаборатория знаний, 2008	ЭБС «Университетская библиотека онлайн»
4	Нестеров С. А.	Интеллектуальный анализ данных средствами MS SQL Server 2008: учебное пособие	Москва: Национальный Открытый Университет «ИНТУИТ», 2016	ЭБС «Университетская библиотека онлайн»
5	Жуковский О. И.	Информационные технологии и анализ данных: учебное пособие	Томск: Эль Контент, 2014	ЭБС «Университетская библиотека онлайн»
6	Байдаков А. Н., Звягинцева О. С., Назаренко А. В., Запорожец Д. В., Бабкина О. Н.	Моделирование бизнес-процессов: учебное пособие	Ставрополь: Ставропольский государственный аграрный университет (СтГАУ), 2017	ЭБС «Университетская библиотека онлайн»
7	Пальмов, С. В.	Интеллектуальный анализ данных: учебное пособие	Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	ЭБС «IPR SMART»
8	Мыльников Л. А., Краузе Б., Кютц М., Баде К., Шмидт И. А.	Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография	Москва: Библио-Глобус, 2017	ЭБС «Университетская библиотека онлайн»
9	Федин, Ф. О., Федин, Ф. Ф.	Анализ данных. Часть 2. Инструменты Data Mining: учебное пособие	Москва: Московский городской педагогический университет, 2012	ЭБС «IPR SMART»
10	Воронов, В. И., Воронова, Л. И., Усачев, В. А.	Data Mining - технологии обработки больших данных: учебное пособие	Москва: Московский технический университет связи и информатики, 2018	ЭБС «IPR SMART»

5.2. Профессиональные базы данных и информационные справочные системы

Информационная справочная система "Консультант Плюс"
База данных Федеральной службы государственной статистики (Росстат) https://rosstat.gov.ru/
База данных Единая межведомственная информационно-статистическая система https://fedstat.ru/
База данных Всероссийского центра исследования общественного мнения (ВЦИОМ) https://wciom.ru/
База данных Центрального банка Российской Федерации (Банка России) https://www.cbr.ru/
База данных Национального агентства финансовых исследований https://nafi.ru/
База данных Фонда общественного мнения https://fom.ru/
База данных Machine learning and data science community https://www.kaggle.com/

5.3. Перечень программного обеспечения

Операционная система РЕД ОС
Офисный пакет LibreOffice (кроссплатформенное свободно распространяемое программное обеспечение)
Прикладной программный пакет Gretl

5.4. Учебно-методические материалы для обучающихся с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:
- столы, стулья;
- персональный компьютер / ноутбук (переносной);
- проектор;
- экран / интерактивная доска.

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет, и/или в специализированных лабораториях, предусмотренных образовательной программой.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

Приложение 1

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

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

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-4: Способен формировать источники больших данных			
Знать - основные источники больших данных сети Интернет	Называет основные источники больших данных; формулирует ответы на вопросы об использовании различных источников больших данных, в том числе сети Интернет, в процессе исследования	Полнота и содержательность ответа; умение приводить примеры; умение отстаивать свою позицию; умение пользоваться дополнительной литературой при подготовке к занятиям; соответствие ответов материалам лекции и учебной литературы, сведениям из информационных ресурсов Интернет	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)
Уметь - подбирать, анализировать и обобщать большие данные из Интернет источников	Использует различные базы данных в процессе выполнения задания к лабораторной работе; производит обобщение и анализ больших данных, извлеченных из сети Интернет в процессе выполнения задания к лабораторной работе; использует современные информационно-коммуникационные технологии и глобальные информационные ресурсы в процессе выполнения задания к лабораторной работе	Полнота и содержательность выполненной лабораторной работы; умение отстаивать свою позицию; обоснованность обращения к базам интернет-данных; целенаправленность поиска и отбора информации из интернет-источников	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)
Владеть - навыками формирования источников больших данных, извлеченных из сети Интернет	Осуществляет формирование источников больших данных из сети Интернет в процессе	Полнота и содержательность выполненной лабораторной работы, глубина анализа; использование различных источников информации Интернет ресурсов, в том	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным

	выполнения задания к лабораторной работе; использует современные пакеты прикладных программ для проведения расчетов в ходе выполнения задания к лабораторной работе;	числе на иностранном языке в целях осуществления определенных этапов анализа данных; умение пользоваться дополнительной литературой, в том числе на иностранном языке при подготовке к занятиям; соответствие отчета по лабораторной работе предъявляемым требованиям.	работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)
ПК-6: Способен осуществлять поиск статистической информации, ее первичную обработку и подготовку для проведения аналитических исследований, в том числе с использованием технологий больших данных			
Знать - основные этапы проведения статистического наблюдения, как основы получения информации для анализа больших данных, извлеченных из сети Интернет	Называет основные этапы проведения статистического наблюдения. Отвечает на вопросы об особенностях сбора информации для осуществления анализа данных	Полнота и содержательность ответа; умение приводить примеры; умение отстаивать свою позицию; умение пользоваться дополнительной литературой при подготовке к занятиям; соответствие ответов материалам лекции и учебной литературы, сведениям из информационных ресурсов Интернет	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)
Уметь - осуществлять анализ результатов проведенного аналитического исследования с использованием технологий больших данных	Применяет основные положения выборочного метода в процессе осуществления ИАД по теме задания к лабораторной работе; производит обработку, анализ, оценивание и интерпретацию результатов проведенного анализа больших данных из сети Интернет в ходе выполнения задания к лабораторной работе; Использует современные пакеты прикладных программ для проведения расчетов к заданию по лабораторной работе	Полнота и содержательность выполненной лабораторной работы; умение отстаивать свою позицию; обоснованность обращения к базам данных интернет ресурсов; целенаправленность поиска и отбора информации из интернет ресурсов; использование различных методов и приемов прикладной и математической статистики в процессе осуществления интеллектуального анализа данных, извлеченных из сети Интернет	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)

Владеть - навыками подготовки статистического инструментария в целях получения первичной информации из сети Интернет для осуществления обработки и анализа данных	Разрабатывает макеты анкет, опросных листов и другого инструментария в целях получения первичной информации для осуществления интеллектуального анализа данных по теме задания к лабораторной работе; составляет отчет по лабораторной работе	Полнота и содержательность выполненной лабораторной работы, глубина анализа; использование системного подхода в процессе осуществления интеллектуального анализа данных, извлеченных из Интернет ресурсов; умение пользоваться дополнительной литературой, в том числе на иностранном языке при подготовке к занятиям; соответствие отчета по лабораторной работе предъявляемым требованиям.	6 семестр: зачетные задания (1-17), опрос (1-39), тесты (1-10), задания к лабораторным работам (1-5); 7 семестр: экзаменационные билеты (1-12), опрос (1-24), тесты (1-10), задания к лабораторным работам (1-4)

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале:

84-100 баллов (оценка «отлично»)
67-83 баллов (оценка «хорошо»)
50-66 баллов (оценка «удовлетворительно»)
0-49 баллов (оценка «неудовлетворительно»)

50-100 баллов (зачтено)
0-49 баллов (не зачтено)

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

6 семестр

ЗАЧЕТНЫЕ ЗАДАНИЯ

Зачетное задание 1

1. Эволюция методов обработки и анализа данных из Интернет ресурсов
2. Схема выполнения задач поиска информации универсальными поисковыми системами

Задача **.** По выборочным данным рассчитаны описательные статистики и оценки параметров модели парной регрессии:

$\tilde{y} = 2{,}4,\ \tilde{x} = 3{,}6,\ \sigma_y = 1{,}07,\ \sigma_x = 1{,}51,\ n = 10.$

$$\hat{y}_i = 0{,}176 + 0{,}618\, x_i$$

$$(0{,}491)\quad (0{,}128)$$

В скобках – стандартные ошибки.

а) Вычислите значение *t*-критерия Стьюдента для b_1. Можете ли Вы утверждать, что он статистически значим на 5%-ном уровне?

б) Вычислите коэффициент вариации для *x* и *y*, дайте оценку характера вариации.

Зачетное задание 2

1. Понятие парсинга открытых данных.
2. Поиск семантических связей в тексте

Задача. Анализ зависимости дохода от количества часов, затраченного руководством фирмы на разработку проектов, привел к следующему уравнению:

Доход = -$957 + $85*количество часов.

а) В соответствии с оценкой зависимости укажите, каким был бы доход (или убытки), если бы на планирование вообще не тратилось время?

б) Насколько в среднем увеличиваются доходы от проектов при увеличении затраченного на планирование времени на 10 часов?

в) Найдите точку самоокупаемости, представляющую собой количество часов, при которых оцениваемая величина дохода равна нулю.

Зачетное задание 3

1. Базы и витрины данных
2. Машинное обучение для поиска по запросу

Задача. По выборочным данным рассчитаны коэффициент детерминации и оценки параметров модели парной регрессии:

$\tilde{y} = 2{,}4,\ \tilde{x} = 3{,}6,\ R^2 = 0{,}717$.

$\hat{y}_i = b_0 + 0{,}618\, x_i$

а) Определите значение b_0.

б) Определите значение коэффициента корреляции r между x и y и его знак.

Зачетное задание 4

1. Интеллектуальный веб-поиск
2. История развития универсальных поисковых систем. Характеристики и возможности универсальных поисковых систем.

Задача. Зависимость уровня дохода (y) (тыс. руб.) от пола (x) описывается следующим уравнением:

$$\hat{y} = 2{,}0 + 0{,}3x.$$

$x = 0$, если работник – женщина,
$x = 1$, если работник – мужчина.

$R^2 = 0{,}24$, $n = 18$.

а) Дайте интерпретацию оценок параметров модели.

б) Какая часть вариации дохода объясняется полом работников? Какая часть объясняется действием других факторов?

в) Проверьте значимость уравнения регрессии.

Зачетное задание 5

1. Сравнительный анализ поисковых систем. Перспективные направления развития универсальных поисковых систем.
2. Методы поиска аудио и видео.

Задача. При построении уравнения множественной регрессии по 20 наблюдениям получены следующие данные:

Переменные	Парные коэффициенты корреляции
Y	$r_{x1,x2} = -0{,}116$
x_1	$r_{y,x1} = 0{,}84$
x_2	$r_{y,x2} = -0{,}21$

а) Проверьте значимость парного коэффициента корреляции между объясняющими переменными.

б) Рассчитайте частные коэффициенты корреляции.

в) Сделайте вывод относительно наличия мультиколлинеарности.

Зачетное задание 6

1. Понятие и структура информационного пространства. Элементы структуры информационного пространства.

2. Булев поиск

Задача. По выборочным данным рассчитаны оценки модели парной регрессии:

SSR = 7,782, SSE = 2,618.

$\hat{y}_i = 0{,}176 + 0{,}618\,x_i$

а) Вычислите значение R^2.

б) Определите значение коэффициента корреляции *r* между *x* и *y* и его знак.

Зачетное задание 7

1. Основные свойства открытых систем.

2. Области применения инструментов парсинга: сбор данных для исследования социально-экономических явлений и процессов, извлечение контактной информации; поиск работы или сотрудников, мониторинг цен

Задача. Зависимость объема продаж от численности населения на соответствующей территории, определяется следующим уравнением:

Ожидаемый объем продаж = $1 371 744 + $0,23675045*численность населения.

R^2=0,37.

а) Дайте интерпретацию коэффициента регрессии.

б) Какая часть вариации объема продаж объясняется численностью населения? Какая часть объясняется действием других факторов?

Зачетное задание 8

1. Концепции организации хранения данных.
2. Автоматическое реферирование.

Задача. По выборочным данным рассчитаны оценки параметров модели динамики прибыли (млн. руб.):

$\hat{y}_t = 236 \cdot 0{,}9^t$.

S_{b0} =0,1, S_{b1} =0,3, n = 10.

а) Дайте интерпретацию оценок параметров модели.

б) С надежностью 0,95 дайте ответ на вопрос, значимы ли оценки параметров модели тренда.

Зачетное задание 9

1. Понятие "показатель". Пространственная интерпретация понятия "показатель". Содержание экономических показателей. Построение системы показателей.
2. Построение и сжатие индекса. Ранжирование документов по степени соответствия запросу.

Задача. По выборочным данным рассчитаны: SSR = 7,782, SSE = 2,618, *n* = 10.

$\hat{y}_i = 0{,}176 + 0{,}618\,x_i$

а) Вычислите значение R^2.

б) Проверьте значимость уравнения регрессии.

Зачетное задание 10

1. Основные инструменты web scraping.
2. Методы интеллектуального веб-поиска.

Задача. Специалист по сельскому хозяйству полагает, что потребление говядины в регионах (y) в тоннах в год зависит от цены говядины (x_1) рублей за килограмм, цены свинины (x_2) рублей за килограмм, цены курятины (x_3) рублей за килограмм и среднедушевых денежных доходов (x_4). Следующая регрессионная модель получена на основе выборки из 30 регионов:

$$\log y = -0.024 - 0.529\log x_1 + 0.217\log x_2 + 0.193\log x_3 + 0.0416\log x_4$$
$$(0.168) \quad (0.103) \quad (0.106) \quad (0.163)$$

$$R^2 = 0.683$$

а) Интерпретируйте коэффициент при *log* x_1.

б) Проверьте на 1% уровне значимости нулевую гипотезу о том, что коэффициент при *log* x_4 в генеральной совокупности равен нулю.

Зачетное задание 11

1. Основные задачи анализа данных, извлеченных из сети Интернет.
2. База метаданных информационного хранилища.

Задача. При оценке параметров модели регрессии по 17 наблюдениям получены следующие данные:

$$\hat{y} = 68{,}236 - 2{,}3x.$$

Коэффициент корреляции между e^2 и $\hat{y}_x$ составил 0,8.

а) Дайте интерпретацию оценок параметров модели.

б) На 5% уровне значимости сделайте вывод о наличии гетероскедастичности.

Зачетное задание 12

1. Этапы и основные методы анализа данных, извлеченных из Интернет источников.
2. Элементы электронного обмена данными.

Задача. Проверить гипотезу о равенстве нулю параметра уравнения регрессии H_0: $\beta_1 = 0$.

$\hat{y} = b_0 + b_1 x$, где $b_1 = 1.71, n = 8, S_{yx} = 0.62, \sigma_x = 0.75$.

Уровень значимости α принять равным 0,01.

Зачетное задание 13

1. Классический поиск. Нечёткий поиск. Web Mining.
2. Методы Web Content Mining.

Задача. . По выборочным данным рассчитаны оценки параметров модели динамики прибыли (млн. руб.):

$$\hat{y}_t = 236 \cdot 0{,}9^t.$$

S_{b0} =0,1, S_{b1} =0,3, n = 10.

а) Дайте интерпретацию оценок параметров модели.

б) С надежностью 0,95 дайте ответ на вопрос, значимы ли оценки параметров модели тренда.

Зачетное задание 14

1. Классификация подходов к поиску информации
2. Алгоритм PageRank.

Задача. Зависимость уровня дохода (*y*) (тыс. руб.) от пола (*x*) описывается следующим уравнением:

$$\hat{y} = 2{,}0 + 0{,}3x.$$

x = 0, если работник – женщина,
x = 1, если работник – мужчина.

$R^2 = 0{,}24$, n = 18.

а) Дайте интерпретацию оценок параметров модели.

б) Какая часть вариации дохода объясняется полом работников? Какая часть объясняется действием других факторов?

в) Проверьте значимость уравнения регрессии.

Зачетное задание 15

1. Векторно-пространственный, вероятностный и языковой поиск.
2. Распознавание образов для поиска изображений.

Задача. По выборочным данным рассчитаны: SSR = 7,782, SSE = 2,618, n = 10.

а) Вычислите значение R2.

б) Проверьте значимость уравнения регрессии.

Зачетное задание 16

1. Методы поиска информации в сети Интернет
2. Оценка достоверности информации. Классификация методов поиска информации

Задача. По выборочным данным рассчитаны описательные статистики и оценки параметров модели парной регрессии:

$\tilde{y}$ = 2,4, $\tilde{x}$ = 3,6, σ_y =1,07, σ_x =1,51, n = 10.

$$\hat{y}_i = 0{,}176 + 0{,}618\,x_i$$
$$(0{,}491)\quad (0{,}128)$$

В скобках – стандартные ошибки.

а) Вычислите значение t-критерия Стьюдента для b_1. Можете ли Вы утверждать, что он статистически значим на 5%-ном уровне?

б) Вычислите коэффициент вариации для x и y, дайте оценку характера вариации.

Зачетное задание 17

1. Алгоритмы предварительной подготовки данных, извлеченных из сети Интернет.
2. Хеширование и деревья поиска.

Задача. По выборочным данным рассчитаны коэффициент детерминации и оценки параметров модели парной регрессии:

= 2,4, = 3,6, .

а) Определите значение b0.

б) Определите значение коэффициента корреляции r между x и y и его знак.

Критерии оценивания для зачета:

Максимальное количество баллов 100. Каждое зачетное задание содержит 2 вопроса и 1 задачу. Ответ на каждый вопрос оценивается отдельно, максимально 30 баллов каждый. Задача оценивается максимально 40 баллов.

Критерии оценивания ответа на отдельный вопрос:

25-30 баллов выставляется студенту, если изложенный материал фактически верен, продемонстрированы глубокие исчерпывающие знания в объеме пройденной программы в соответствии с поставленными программой курса целями и задачами обучения, изложение материала при ответе - грамотное и логически стройное;

20-24 балла выставляется студенту, если продемонстрированы твердые и достаточно полные знания в объеме пройденной программы дисциплины в соответствии с целями обучения; материал изложен достаточно полно с отдельными логическими и стилистическими погрешностями;

15-19 баллов выставляется студенту, если продемонстрированы твердые знания в объеме пройденного курса в соответствие с целями обучения, ответ содержит отдельные ошибки, уверенно исправленные после дополнительных вопросов;
0-14 баллов выставляется студенту, если ответы не связаны с вопросами, допущены грубые ошибки в ответе, продемонстрированы непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Критерии оценивания отдельной задачи:

35-40 баллов выставляется, если задача решена полностью, в представленном решении обоснованно получены правильные ответы, проведен анализ, дана грамотная интерпретация полученных результатов, сделаны выводы.
25-34 балла выставляется, если задача решена полностью, но при анализе и интерпретации полученных результатов допущены незначительные ошибки, выводы – достаточно обоснованы, но неполны.
11-24 балла выставляется, если задача решена частично, анализ и интерпретация полученных результатов не вполне верны, выводы верны частично.
0-10 баллов выставляется, если решение неверно или отсутствует.

Итоговый результат формируется из суммы набранных баллов за выполнение зачетного задания и соответствует шкале:

- 50-100 баллов (зачтено);
- 0-49 баллов (не зачтено).

Опрос

Вопросы для проведения опроса:

1. Определите основные задачи анализа данных, извлеченных из сети Интернет.
2. Раскройте этапы и основные методы анализа данных, извлеченных из Интернет источников.
3. Проанализируйте эволюцию методов обработки и анализа данных из Интернет ресурсов.
4. Раскройте суть и основные этапы алгоритмов предварительной подготовки данных, извлеченных из сети Интернет.
5. Дайте понятие «Базы данных» и «Витрины данных». Назовите основные октрытые базы и витрины данных в сети Интернет. В чем их преимущества и недостатки?
6. Опишите механизм осуществления классического поиска данных в сети Интернет.
7. Опишите механизм осуществления нечёткого поиска данных в сети Интернет и Web Mining.
8. Опишите процедуру анализ изображений, аудио и видео сети Интернет.
9. Выявите особенности и этапы осуществления интеллектуального веб-поиска.
10. Выделите классификацию подходов к поиску информации в сети Интернет
11. Раскройте понятие парсинга открытых данных. Выделите нормативно-правовую базу парсинга открытых данных.
12. Какие основные инструменты web scraping вы знаете?
13. Каковы области применения инструментов парсинга?
14. Какие методы поиска информации в сети Интернет вы знаете?
15. В чем особенность осуществления Булева поиск информации в сети Интернет?
16. В чем особенность векторно-пространственного, вероятностного и языкового поиск информации в сети Интернет?
17. Опишите методику хеширования и построения деревьев поиска.
18. Какие методы Web Content Mining вы знаете?
19. В чем суть машинного обучения для поиска по запросу?
20. В чем суть латентно-семантического индексирования?
21. Выделите основные методики распознавания образов для поиска изображений в сети Интернет
22. Выделите основные методы поиска аудио и видео в сети Интернет
23. Выделите основные методы интеллектуального веб-поиска.
24. Опишите алгоритм оценки достоверности информации и классификации методов поиска информации в сети интернет

25. Опишите процесс построения и сжатия индекса; ранжирования документов по степени соответствия запросу.
26. Опишите алгоритм PageRank.
27. Опишите алгоритмы классификации и кластеризации.
28. Опишите процесс автоматическое реферирования в сети Интернет
29. Раскройте алгоритм поиска семантических связей в тексте в сети Интернет
30. Как происходит оценка смыслового подобия?
31. Как происходит поиск ответов на вопрос в сети Интернет?
32. Опишите основные свойства открытых систем.
33. Какие элементы электронного обмена данными вы знаете?
34. Как осуществляется процесс оценки и повышения качества данных и приведения их к одной структуре?
35. Рассмотрите историю развития универсальных поисковых систем.
36. Каковы характеристики и возможности универсальных поисковых систем?
37. Проведите сравнительный анализ поисковых систем.
38. Выявите перспективные направления развития универсальных поисковых систем.
39. Опишите схему выполнения задач поиска информации универсальными поисковыми системами

Критерии оценивания

Максимальный балл – 10. В течение семестра обучающийся отвечает максимум на 10 вопросов. Ответ на каждый вопрос оценивается максимум в 1 балл.

Критерии оценивания одного вопроса:

0,84 – 1 балла выставляется студенту, если изложенный материал фактически верен, продемонстрированы глубокие исчерпывающие знания в объеме пройденной программы в соответствии с поставленными программой курса целями и задачами обучения, изложение материала при ответе - грамотное и логически стройное;

0,67 – 0,83 балла выставляется студенту, если продемонстрированы твердые и достаточно полные знания в объеме пройденной программы дисциплины в соответствии с целями обучения; материал изложен достаточно полно с отдельными логическими и стилистическими погрешностями;

0,5 – 0,66 балла выставляется студенту, если продемонстрированы твердые знания в объеме пройденного курса в соответствие с целями обучения, ответ содержит отдельные ошибки, уверенно исправленные после дополнительных вопросов;

0-0,49 балла выставляется студенту, если ответы не связаны с вопросами, допущены грубые ошибки в ответе, продемонстрированы непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Тесты

1. Ошибки репрезентативности (представительности) - это:

a) разность между характеристиками выборочной совокупности и генеральной совокупностей
b) разность между истинными и зарегистрированными значениями признака
c) среднее квадратическое отклонение возможных значений выборочной характеристики от характеристики генеральной совокупности, взвешенных по вероятностям их наступления
d) сумму отклонений возможных значений выборочной средней от генеральной средней, взвешенных по вероятностям их наступления

2. Систематические ошибки выборки возникают вследствие:

a) ошибок печати
b) нарушения принципа случайности отбора
c) ошибок в вычислении предельной ошибки выборки
d) слишком большого объёма выборки

3. Предельная ошибка выборки позволяет определять:

a) надёжность результатов, полученных по данным выборки
b) предельные значения характеристик генеральной совокупности при заданной доверительной вероятности

c) вероятность расхождения выборочных и генеральных характеристик
d) минимально возможные расхождения выборочных и генеральных характеристик

4. Стандартная ошибка выборки представляет собой

a) среднее квадратическое отклонение возможных значений выборочной характеристики от характеристики генеральной совокупности, взвешенных по вероятностям их наступления
b) сумму отклонений возможных значений выборочной средней от генеральной средней, взвешенных по вероятностям их наступления
c) отклонение генеральной средней от предельной ошибки выборки
d) отклонение выборочной средней от предельной ошибки выборки

5. Предельная ошибка выборки равна

a) сумме стандартной ошибки и величины кратности ошибки
b) частному от деления величины кратности ошибки и стандартной ошибки
c) разности стандартной ошибки и величины кратности ошибки
d) t-кратному числу стандартных ошибок выборки

6. Статистическая оценка является несмещённой, если:

a) её дисперсия достаточно велика
b) её математическое ожидание равно оцениваемому параметру генеральной совокупности
c) её ожидаемое значение стремится к параметру генеральной совокупности
d) она имеет небольшую дисперсию

7.Задача классификации сводится к …

а) нахождения частых зависимостей между объектами или событиями;
б) определения класса объекта по его характеристиками;
в) определение по известным характеристиками объекта значение некоторого его параметра;
г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

8. Задача регрессии сводится к …

а) нахождения частых зависимостей между объектами или событиями;
б) определения класса объекта по его характеристиками;
в) определение по известным характеристиками объекта значение некоторого его параметра;
г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

9. Задача кластеризации заключается в …

а) нахождения частых зависимостей между объектами или событиями;
б) определения класса объекта по его характеристиками;
в) определение по известным характеристиками объекта значение некоторого его параметра;
г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

10. Целью поиска ассоциативных правил является …

а) нахождения частых зависимостей между объектами или событиями;
б) определения класса объекта по его характеристиками;
в) определение по известным характеристиками объекта значение некоторого его параметра;
г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

Критерии оценивания:

Максимальная сумма по итогам решения тестов - 10 баллов
Каждый вопрос оценивается в 1 балл, если ответ верный и 0 баллов, если неверный

Задания к лабораторным работам

Задание к лабораторной работе 1. Обзор основных открытых баз и витрин данных в целях осуществления ИАД.

1. Рассмотреть основные базы открытых данных: Федеральной службы государственной статистики (Росстат), органов исполнительной власти на разных уровнях территориальной агрегации, Единой межведомственной информационно-статистической системы, Всероссийского центра исследования общественного мнения (ВЦИОМ), Центрального банка РФ, Национального агентства финансовых исследований, Фонда общественного мнения, портала открытых данных РФ (Data.gov.ru), kaggle.com и т.п.
2. Рассмотреть структуру и наполнение отдельных видов баз данных;

3. Выявить преимущества отдельных видов баз данных для осуществления интеллектуального анализа данных конкретных профессиональных задач.
4. Подготовить отчет по лабораторной работе 1.

Задание к лабораторной работе 2. Извлечение и анализ данных из открытых источников.

1. Осуществить выбор темы исследования, по которой будет производится извлечение и анализ данных из открытых источников;
2. Произвести поиск необходимой информации в локальных поисковых системах;
3. Осуществить поиск необходимой информации в открытых базах данных;
4. Произвести обработку первичной информации, полученной на этапах 2 и 3 лабораторной работы;
5. Осуществить визуализацию данных, по рассматриваемой тематике собранных и обработанных на предыдущих этапах;
6. Произвести анализ и оценку данных по рассматриваемой тематике, полученных на предыдущих этапах лабораторной работы.
7. Подготовить отчет по лабораторной работе 2.

Задание к лабораторной работе 3. Классификация и кластеризация данных

1. Осуществить выбор темы исследования, по которой будет производится классификация и кластеризация данных
2. Осуществить выбор данных для осуществления классификации и кластерного анализа
3. Рассмотреть основные методы кластеризации данных
4. Осуществить классификацию выбранных данных
5. Выполнить кластеризацию объектов с помощью иерархических агломеративных методов, используя различные меры расстояния между объектами:
6. Выполнить кластеризацию объектов на основе использования качественных признаков.
7. Провести кластеризацию объектов, используя метод К-средних и количественные показатели в качестве исходных данных.
8. Построить правило отнесения объектов к одному из выделенных классов (кластеров) на основе методов дискриминантного анализа.
9. Сформулировать выводы по проведенному кластерному анализу данных
10. Подготовить отчета по лабораторной работе 3.

Задание к лабораторной работе 4. Методы графического представления и визуализации данных

1. Осуществить выбор темы исследования, по которой будет производится графическое представление и визуализация данных
2. Произвести выбор данных для их дальнейшего графического представления и визуализации
3. Рассмотреть основные способы графического представления и визуализации данных.
4. Осуществить выбор способов графического представления и визуализации данных для решения конкретной профессиональной задачи на основе выбранных данных
5. Осуществить графическое представление и визуализацию выбранных данных
6. Подготовить отчет по лабораторной работе 4.

Задание к лабораторной работе 5. Реализация и интерпретация результатов многомерного статистического анализа.

1. Осуществить постановку цели, задач исследования; выбор объекта исследования.
2. Произвести выбор данных для реализации многомерного статистического анализа
3. Осуществить корреляционно-регрессионный анализ на основе выбранных данных
4. Осуществить дискриминантный анализ с учетом цели и задач исследования
5. Произвести факторный анализ на основе выбранных данных
6. Сформулировать основные выводы на основе проведенного комплексного многомерного анализа

7. Подготовить отчет по лабораторной работе 5.

Критерии оценивания:
Каждое лабораторное задание оценивается отдельно по 16 баллов. Максимальная сумма баллов – 80.

Критерии оценивания отдельного лабораторного задания:
14-16 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет;
11-13 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет; дана содержательная интерпретация полученных при решении задач результатов; материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;
8-10 баллов выставляется, если обучающийся: выполнил работу в полном объеме с соблюдением необходимой последовательности; грамотно оформил представленный отчет; дана содержательная интерпретация полученных результатов; допускаются отдельные логические и стилистические погрешности; обучающийся может испытывать некоторые затруднения в формулировке суждений;
0-7 баллов выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

7 семестр

Экзаменационные билеты

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №1

1. Эволюция методов обработки и анализа данных из Интернет ресурсов.
2. Специализированные средства создания и поддержки анализа данных Интернет ресурсов, их характеристика.

Задача. Опишите процесс проведения исследования в рамках анализа динамики промышленного производства в долгосрочном периоде. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №2

1. Возможности автоматизации отдельных этапов анализа данных, извлеченных из Интернет источников.
2. Применение данных сети Интернет в маркетинговой деятельности и в сфере логистики.

Задача. Опишите процесс проведения исследования в рамках анализа динамики потребительского спроса в долгосрочном периоде. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №3

1. Осуществление импорта/экспорта данных интернет ресурсов. Визуализация результатов
2. Осуществление анализа цепочек и издержек логистических процессов с помощью ресурсов сети Интернет

Задача. Опишите процесс проведения исследования в рамках анализа государственного регулирования экономического роста. Укажите информационные ресурсы и базы данных,

необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №4

1. Этапы и основные методы анализа данных, извлеченных из Интернет источников.
2. Осуществление финансового и инвестиционного анализа данных сети Интернет.

Задача. Опишите процесс проведения исследования в рамках оценка доли фирмы на рынке и ее объема продаж. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №5

1. Аналитические инструментальные средства пакетов прикладных программ широкого применения LibreOffice.
2. Блоки интегрированных экономических информационных систем.

Задача. Опишите процесс проведения исследования в рамках анализа функционирования отдельных подсистем предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №6

1. Классификация подходов к поиску информации
2. Осуществление понижения размерности, сглаживания аномалий, фильтрации, группировки данных, извлеченных из сети Интернет

Задача. Опишите процесс проведения исследования в рамках анализа конкурентных преимуществ предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №7

1. Классификация IT-анализа по режиму и темпу.
2. Применение данных, извлеченных из сети Интернет для решения бизнес-задач. Отбор данных для анализа. Предварительная подготовка данных.

Задача. Опишите процесс проведения исследования в рамках анализа маркетинговой стратегии предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №8

1. Задачи и содержание оперативного (OLAP) анализа. Типы многомерных OLAP систем.
2. Осуществление анализа ставки дисконтирования с поправкой на риск с помощью ресурсов сети Интернет

Задача. Опишите процесс проведения исследования в рамках анализа потоков платежей предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №9

1. Аналитические платформы Data Mining, используемые для анализа данных, извлеченных из Интернет источников: основные возможности, сфера применения, этапы решения задач.
2. Осуществление анализа разрыва; портфолио-анализа; анализа маржинальной прибыли предприятия с помощью ресурсов сети Интернет

Задача. Опишите процесс проведения исследования в рамках анализа инвестиционной привлекательности отрасли или региона. Укажите информационные ресурсы и базы данных,

необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №10

1. Подходы к выполнению анализа данных Интернет ресурсов средствами информационных технологий (IT-анализа).
2. Осуществление анализа ставки дисконтирования с поправкой на риск с помощью Интернет ресурсов

Задача. Опишите процесс проведения исследования в рамках анализа стратегической позиции предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №11

1. Типы инструментальных средств создания и поддержки анализа данных Интернет ресурсов.
2. Осуществление анализа потоков платежей. Финансовая "паутина". Функционально-стоимостной анализ.

Задача. Опишите процесс проведения исследования в рамках проектирования политики в сфере логистики предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ №12

1. Назначение инструментальных средств, их классификация.
2. Анализ обеспечения ресурсами. Анализ возможных прерываний бизнес-процессов. Определение верхних границ цен. Анализ стратегической позиции предприятия

Задача. Опишите процесс проведения исследования в рамках анализа эффективности деятельности предприятия. Укажите информационные ресурсы и базы данных, необходимые для осуществления данного анализа. Опишите процедуру поиска и извлечения данных из сети Интернет в целях осуществления исследования по данной тематике.

Критерии оценивания для экзамена:

Максимальное количество баллов 100. Каждый экзаменационный билет содержит 2 вопроса и 1 задачу. Ответ на каждый вопрос оценивается отдельно, максимально 30 баллов каждый. Задача оценивается максимально в 40 баллов.

Критерии оценивания ответа на отдельный вопрос:

25-30 баллов выставляется студенту, если изложенный материал фактически верен, продемонстрированы глубокие исчерпывающие знания в объеме пройденной программы в соответствии с поставленными программой курса целями и задачами обучения, изложение материала при ответе - грамотное и логически стройное;

20-24 баллов выставляется студенту, если продемонстрированы твердые и достаточно полные знания в объеме пройденной программы дисциплины в соответствии с целями обучения; материал изложен достаточно полно с отдельными логическими и стилистическими погрешностями;

15-19 баллов выставляется студенту, если продемонстрированы твердые знания в объеме пройденного курса в соответствие с целями обучения, ответ содержит отдельные ошибки, уверенно исправленные после дополнительных вопросов;

0 - 14 балла выставляется студенту, если ответы не связаны с вопросами, допущены грубые ошибки в ответе, продемонстрированы непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Критерии оценивания отдельной задачи:

35-40 баллов выставляется, если задача решена полностью, в представленном решении обоснованно получены правильные ответы, проведен анализ, дана грамотная интерпретация полученных результатов, сделаны выводы.

25-34 балла выставляется, если задача решена полностью, но при анализе и интерпретации полученных результатов допущены незначительные ошибки, выводы – достаточно обоснованы, но неполны.
11-24 балла выставляется, если задача решена частично, анализ и интерпретация полученных результатов не вполне верны, выводы верны частично.
0-10 баллов выставляется, если решение неверно или отсутствует.

Итоговый результат формируется из суммы набранных баллов за выполнение экзаменационного билета и соответствует шкале:

- 84-100 баллов (оценка «отлично»)
- 67-83 баллов (оценка «хорошо»)
- 50-66 баллов (оценка «удовлетворительно»)
- 0-49 баллов (оценка «неудовлетворительно»)

Опрос

Вопросы для проведения опроса:

1. Рассмотрите основные возможности, сферу применения, этапы решения задач аналитических платформ Data Mining, используемые для анализа данных, извлеченных из Интернет источников
2. Опишите алгоритм импорта/экспорта данных. Как происходит визуализация результатов?
3. Каковы возможности автоматизации отдельных этапов анализа данных, извлеченных из Интернет источников?
4. Выделите основные подходы к выполнению анализа данных Интернет ресурсов средствами информационных технологий (IT-анализа).
5. Рассмотрите классификацию IT-анализа по режиму и темпу.
6. Определите задачи и содержание оперативного (OLAP) анализа.
7. Выделите типы многомерных OLAP систем.
8. Выделите типы инструментальных средств создания и поддержки анализа данных Интернет ресурсов.
9. Определите назначение инструментальных средств, их классификация.
10. Рассмотрите аналитические инструментальные средства пакетов прикладных программ широкого применения Libre Office.
11. Рассмотрите специализированные средства создания и поддержки анализа данных Интернет ресурсов.
12. Выделите блоки интегрированных экономических информационных систем.
13. Какова область применения данных, извлеченных из сети Интернет для решения бизнес-задач? Как осуществляется отбор данных для анализа? Выделите этапы предварительной подготовки данных.
14. Раскройте суть методик понижения размерности, сглаживания аномалий, фильтрации, группировки при решении бизнес-задач
15. Раскройте суть и основные этапы осуществления анализа обеспечения ресурсами предприятия на основе данных сети Интернет.
16. Раскройте суть и основные этапы осуществления анализа возможных прерываний бизнес-процессов на основе данных сети Интернет.
17. Как происходит осуществление методики определения верхних границ цен?
18. Как произвести анализ стратегической позиции предприятия на основе данных интернет ресурсов?
19. Опишите области применение данных сети Интернет в маркетинговой деятельности и в сфере логистики.
20. Как происходит анализ разрыва; портфолио-анализ; анализ маржинальной прибыли организации на основе данных сети интернет?
21. Выделите основные этапы анализа цепочек и издержек логистических процессов с помощью интернет ресурсов
22. Выявите основные этапы и методики осуществления финансового и инвестиционного анализ данных сети Интернет.

23. Как осуществляется анализ потоков платежей; анализ «Финансовая "паутина"»; функционально-стоимостной анализ на основе интернет ресурсов?
24. Определите суть и этапы осуществления анализа ставки дисконтирования с поправкой на риск с помощью интернет ресурсов.

Критерии оценивания

Максимальный балл – 10.
В течение семестра обучающийся отвечает максимум на 10 вопросов. Ответ на каждый вопрос оценивается максимум в 1 балл.

Критерии оценивания одного вопроса:

0,84 – 1 балла выставляется студенту, если изложенный материал фактически верен, продемонстрированы глубокие исчерпывающие знания в объеме пройденной программы в соответствии с поставленными программой курса целями и задачами обучения, изложение материала при ответе - грамотное и логически стройное;

0,67 – 0,83 балла выставляется студенту, если продемонстрированы твердые и достаточно полные знания в объеме пройденной программы дисциплины в соответствии с целями обучения; материал изложен достаточно полно с отдельными логическими и стилистическими погрешностями;

0,5 – 0,66 балла выставляется студенту, если продемонстрированы твердые знания в объеме пройденного курса в соответствие с целями обучения, ответ содержит отдельные ошибки, уверенно исправленные после дополнительных вопросов;

0-0,49 балла выставляется студенту, если ответы не связаны с вопросами, допущены грубые ошибки в ответе, продемонстрированы непонимание сущности излагаемого вопроса, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Тесты

1. Корреляционная связь - это:
a) вид причинной зависимости, проявляющейся не в каждом отдельном случае, а в общем, в среднем, при большом числе наблюдений
b) зависимость среднего значения результативного признака от изменения факторного признака
c) вид причинной зависимости, при которой определенному значению факторного признака соответствует одно или несколько точно заданных значений результативного признака
d) причинно-следственная связь явлений и процессов, когда изменение одного из них -причины ведет к изменению другого – следствия

2. До предполагаемых моделей относятся такие модели данных:
а) модели классификации и последовательностей;
б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
в) классификации, кластеризации, исключений, итоговые и ассоциации;
г) модели классификации, последовательностей и исключений.

3. В описательных моделей относятся следующие модели данных:
а) модели классификации и последовательностей;
б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
в) классификации, кластеризации, исключений, итоговые и ассоциации;
г) модели классификации, последовательностей и исключений.

4. Модели классификации описывают …
а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
г) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.

5. Коэффициент корреляции Пирсона характеризует
а) тесноту и направление связи между двумя коррелируемыми признаками в случае наличия между

ними линейной зависимости
b) тесноту связи наличия нелинейной зависимости между двумя признаками
c) тесноту связи двумя качественных признаков
d) тесноту связи между альтернативными и варьирующими признаками

6. Коэффициент корреляции Пирсона принимает значения:

a) от +1 до -1
b) от 0 до +1
c) от-1 до 0
d) только положительные значения

7. При каком значении коэффициента корреляции Пирсона связь между результативным и факторными признаками можно считать достаточно сильной:

a) 1,5
b) -0,5
c) -1,2
d) 0,8

8. Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции Пирсона:

a) не изменится
b) увеличится (уменьшится) на то же число
c) уменьшится (увеличится) в то же число раз
d) увеличится (уменьшится) на то же число или в то же число раз

9. Как влияет изменение числа наблюдений на надежность коэффициента корреляции?

a) с уменьшением числа наблюдений надежность коэффициента корреляции падает
b) с увеличением числа наблюдений надежность коэффициента корреляции падает
c) с уменьшением числа наблюдений надежность коэффициента корреляции растет
d) при прочих равных условиях число наблюдений не влияет на надежность коэффициента корреляции

10. При выявлении статистической зависимости по данным аналитической группировки в качестве меры степени тесноты связи может быть использовано:

a) эмпирическое корреляционное отношение
b) теоретическое корреляционное отношение
c) коэффициент детерминации
d) коэффициент конкордации

Критерии оценивания:

Максимальная сумма по итогам решения тестов - 10 баллов
Каждый вопрос оценивается в 1 балл, если ответ верный и 0 баллов, если неверный

Задания к лабораторным работам

Задание к лабораторной работе 1. Реализация и интерпретация процедуры интеллектуального анализа временных рядов и прогнозирования.

1. Осуществить постановку цели, задач исследования; выбор объекта исследования.
2. Произвести выбор данных для реализации процедуры интеллектуального анализа временных рядов и прогнозирования
3. Предварительный анализ исследуемого процесса: выдвижение гипотез о возможном развитии исследуемого объекта в ближайшем будущем; выявление и общий анализ факторов, благоприятствующих или препятствующих развитию объекта; анализ вопроса об инерционности развития объекта; установление или уточнение экономического показателя, наиболее полно характеризующего объект прогнозного исследования; определение базы прогноза (периода времени в прошлом, на базе которого будет собираться информация о прогнозном объекте); определение допустимого периода упреждения; определение шага базы прогноза (год, месяц, неделя, день и т.д.); определение содержания уровней временного ряда (абсолютные, относительные, средние и т.д.).
4. Построение модели тренда: анализ данных и выбор формы модели (Визуальный анализ. Построение и анализ графика. Вывод на основе визуального анализа)

5. Аналитические методы обоснования выбора модели тренда: а) сглаживание временного ряда, применение скользящих средних; б) расчет конечных разностей (для полиномов); в) использование аналитических критериев; общий вывод.

6. Количественная оценка параметров модели тренда на основе метода наименьших квадратов.

7. Проверка качества модели

8. Прогнозирование экономического показателя на основе модели тренда, путем экстраполяции.

9. Подготовка отчета по лабораторной работе 1.

Задание к лабораторной работе 2. Реализация и интерпретация результатов интеллектуального анализа данных для решения основных бизнес задач

1. Осуществить постановку цели, задач интеллектуального анализа определенного бизнес-процесса и бизнес-задачи

2. Произвести выбор источников информации для осуществления интеллектуального анализа бизнес-задачи

3. Осуществить выбор методов для проведения интеллектуального анализа бизнес-задачи с учетом заявленной цели

4. Осуществить интеллектуальный анализ бизнес-процесса с учетом методов, выбранных в пункте 3.

5. Сформулировать основные выводы и предложения по результатам проведенного интеллектуального анализа бизнес-задачи

6. Подготовить отчет по лабораторной работе 2.

Задание к лабораторной работе 3. Реализация и интерпретация результатов интеллектуального анализа данных в маркетинговой деятельности и в сфере логистики

1. Осуществить постановку цели, задач интеллектуального анализа определенного направления маркетинговой деятельности или сферы логистики

2. Произвести выбор источников информации для осуществления интеллектуального анализа определенного направления маркетинговой деятельности или сферы логистики

3. Осуществить выбор методов проведения интеллектуального анализа определенного направления маркетинговой деятельности или сферы логистик с учетом заявленной цели

4. Осуществить интеллектуальный анализ определенного направления маркетинговой деятельности или сферы логистики с учетом выбранных методов в п.3.

5. Сформулировать основные выводы и предложения по результатам проведенного интеллектуального анализа определенного направления маркетинговой деятельности или сферы логистики

6. Подготовить отчет по лабораторной работе 3.

Задание к лабораторной работе 4. Реализация и интерпретация результатов финансового и инвестиционного анализа

1. Осуществить постановку цели, задач интеллектуального анализа определенного направления финансовой или инвестиционной деятельности

2. Осуществить выбор источников информации для проведения интеллектуального анализа определенного направления финансовой или инвестиционной деятельности

3. Произвести выбор методов осуществления интеллектуального анализа определенного направления финансовой или инвестиционной деятельности с учетом заявленной цели

4. Осуществить интеллектуальный анализ определенного направления финансовой или инвестиционной деятельности с учетом методов из п.3.

5. Сформулировать основные выводы и предложения по результатам проведенного интеллектуального анализа определенного направления финансовой или инвестиционной деятельности

6. Подготовить отчет по лабораторной работе 4.

Критерии оценивания:
Каждое лабораторное задание оценивается отдельно по 20 баллов каждое. Максимальная сумма баллов – 80.

Критерии оценивания отдельного лабораторного задания:

18 - 20 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет;
14 - 17 баллов выставляется, если обучающийся: выполнил работу в полном объеме, самостоятельно, с соблюдением необходимой последовательности; грамотно оформил представленный отчет; дана содержательная интерпретация полученных при решении задач результатов; материал изложен четко; допускаются отдельные логические и стилистические погрешности, уверенно исправленные после дополнительных вопросов;
11 – 13 выставляется, если обучающийся: выполнил работу в полном объеме с соблюдением необходимой последовательности; грамотно оформил представленный отчет; дана содержательная интерпретация полученных результатов; допускаются отдельные логические и стилистические погрешности; обучающийся может испытывать некоторые затруднения в формулировке суждений;
0 – 10 выставляется, если работа не выполнена или выполнена не в полном объеме; обучающийся практически не владеет теоретическим материалом, допуская грубые ошибки, испытывает затруднения в формулировке собственных суждений, неспособен ответить на дополнительные вопросы.

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

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме зачета, экзамена.

Зачет проводится по расписанию промежуточной аттестации в письменном виде. В зачетном задании – 2 теоретических вопроса и 1 задача. Проверка ответов и объявление результатов производится в день зачета. Результаты аттестации заносятся в ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику, должны ликвидировать задолженность в установленном порядке.

Экзамен проводится по расписанию промежуточной аттестации в письменном виде. В экзаменационном билете – 2 теоретических вопроса и 1 задача. Проверка ответов и объявление результатов производится в день экзамена. Результаты аттестации заносятся в ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику, должны ликвидировать задолженность в установленном порядке.

Приложение 2

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

– лекционные занятия;

– лабораторные занятия.

В ходе лекционных занятий рассматриваются теоретические вопросы и практические примеры реализации методов извлечения и анализа Интернет-данных, даются рекомендации для самостоятельной работы и подготовке к лабораторным занятиям.

В ходе лабораторных занятий развиваются навыки применения методов интеллектуального анализа данных для решения конкретных задач.

При подготовке к лабораторным занятиям студент должен:

– изучить рекомендованную учебную литературу;

– подготовить ответы на все вопросы по изучаемой теме;

В процессе подготовки к лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекционных и лабораторных занятиях, должны быть изучены студентами в ходе самостоятельной работы. Студент должен готовиться к предстоящему лабораторному занятию по всем, обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.

Методические рекомендации по написанию, требования к оформлению отчета по лабораторной работе

Отчет должен содержать следующее: титульный лист; содержание отчета; исходные данные для осуществления расчетов и анализа; основные этапы осуществления лабораторной работы в соответствии с программой; выводы, предложения и критические замечания; приложения.

Требования, предъявляемые к содержанию основных разделов отчета:

- четкость и логическая последовательность изложения материала;
- убедительность аргументации;
- краткость и четкость формулировок, исключающих возможность неоднозначного толкования.

Отчет должен содержать описание действий с иллюстрациями, формулами, результаты выполнения лабораторной работы в тех или иных программных средствах.

Отчет по лабораторной работе выполняется на стандартных листах белой бумаге, на одной стороне, формата А4 (210х297 мм). Текст печатается шрифтом Liberation Serif № 12 через 1 интервал.

Текст работы должен быть аккуратно оформлен и экономически грамотно изложен с учетом требований современной орфографии.

Все листы работы (текстовые, табличные) должны быть выполнены с соблюдением следующих минимальных размеров полей: с левой стороны - 30 мм; правой - не менее 10 мм; сверху - и снизу - 20 мм. При этом текст рамкой не очерчивается.

Страницы в отчете должны иметь сквозную нумерацию.