

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Федеральное государственное бюджетное образовательное учреждение высшего образования

«Ростовский государственный экономический университет (РИНХ)»

Дата подписания: 01.09.2025 22:06:45

Уникальный программный ключ:

c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

УТВЕРЖДАЮ

Начальник

учебно-методического управления

Т.К. Платонова

«20» мая 2025 г.

Рабочая программа дисциплины
Data science в цифровой экономике

Направление подготовки
38.04.08 Финансы и кредит

Направленность (профиль) программы магистратуры
38.04.08.03 Финансовый мониторинг и финансовые рынки

Для набора 2025 года

Квалификация
Магистр

КАФЕДРА Финансовый мониторинг и финансовые рынки**Распределение часов дисциплины по семестрам / курсам**

Курс	1		Итого	
Вид занятий	уп	рп		
Лекции	4	4	4	4
Практические	4	4	4	4
Итого ауд.	8	8	8	8
Контактная работа	8	8	8	8
Сам. работа	91	91	91	91
Часы на контроль	9	9	9	9
Итого	108	108	108	108

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.02.2025 г. протокол № 9.

Программу составил(и): д.э.н., проф., Иванченко И.С.

Зав. кафедрой: д.э.н., профессор Ю.С. Евлахова

Методический совет направления: д.э.н., профессор О.Б. Иванова

Директор института магистратуры: д.э.н., профессор Е.А. Иванова

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Основной целью изучения дисциплины является выработка у студентов компетенций и навыков по работе большими массивами данных, а также овладение методами обработки этих массивов данных и получение агрегированных показателей.
-----	--

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

ПК-5. Способен выявлять и проводить исследование актуальных научных проблем в области финансового мониторинга и финансовых рынков

В результате освоения дисциплины обучающийся должен:

Знать:

правила построения баз данных, а также методики сбора данных, их проверки и фильтрации (соотнесено с индикатором УК-1.1);

основные методы прикладной статистики, необходимых для проведения интеллектуального анализа данных на финансовом рынке (соотнесено с индикатором ПК-5.1);

Уметь:

выбирать модели (алгоритмы) анализа данных; осуществлять подбор параметров модели и алгоритма обучения (соотнесено с индикатором УК-1.2);

осуществлять анализ результатов проведенного исследования на основе методов прикладной статистики (соотнесено с индикатором ПК-5.2)

Владеть:

навыками анализа большого объема информации, выявления скрытых закономерностей, а также прогнозирования динамики временных рядов (соотнесено с индикатором УК-1.3);

навыками освоения основных методов прикладной статистики в процессе осуществления Data Mining (соотнесено с индикатором ПК-5.3).

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Data mining и базы данных

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
1.1	Тема 1.1. "Введение в Data Mining". Определение термина "Data Mining". Задачи "Data Mining": информация и знание; классификация и кластеризация; прогнозирование и визуализация. Сферы применения "Data Mining". Основы анализа данных.	Лекционные занятия	1	2	УК-1 ПК-5
1.2	Тема 1.1. "Введение в Data Mining". Исторический экскурс в обработку больших данных. Обсуждение вопросов, связанных с поиском и обработкой данных, раскрытие термина "Data Mining". Освещение задач "Data Mining". Сферы применения "Data Mining", обсуждение особенностей анализа данных.	Практические занятия	1	2	УК-1 ПК-5
1.3	Тема 1.2. "Общая характеристика методов интеллектуального анализа данных". История развития методов анализа и обработки данных. Переход от статистической обработки к методам OLAP и Data Mining. Принципы, лежащие в основе методов OLAP и Data Mining. Обзор основных задач, решаемых методами Data Mining.	Самостоятельная работа	1	2	УК-1 ПК-5
1.4	Тема 1.2. "Общая характеристика методов интеллектуального анализа данных". «Большие данные» как о социально-экономический феномен. Обсудить определяющие характеристики больших данных. Необходимость высокоскоростной обработки больших данных и получение результатов. Обсуждение причин использования OLAP для обработки запросов.	Самостоятельная работа	1	4	УК-1 ПК-5
1.5	Тема 1.3. "Правила построения баз данных". "Правила построения баз данных". Определение базы данных, история возникновения и развития. Виды баз данных. Классификация баз данных по различным критериям: по модели данных, по среде постоянного хранения, по содержанию, по степени распространенности. Сверхбольшие базы данных. Составление простейших Select-запросов. Формирование базы данных с заданными характеристиками и требуемой информацией.	Самостоятельная работа	1	7	УК-1 ПК-5
1.6	Тема 1.4. "Введение в SQL-запросы". Составление простейших Select-запросов. Формирование базы данных с заданными	Самостоятельная работа	1	2	УК-1 ПК-5

	характеристиками и требуемой информацией.				
1.7	Разработка связанных баз данных и набора SQL-запросов, которые бы помогали собирать информацию с конкретного сегмента финансового рынка для выработки инвестиционных решений.	Самостоятельная работа	1	36	УК-1 ПК-5
Раздел 2. Методы классификации и прогнозирования данных					
№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
2.1	Тема 2.1. "Алгоритмы предварительной обработки данных". Задачи предварительной подготовки данных для анализа. Тестирование данных на выявление или отсутствие в их рядах трендовых зависимостей. Понятие информационной эффективности рынков. Выбор математического инструментария анализа данных в зависимости от степени информационной эффективности рынка. Очистка данных: факторный анализ, корреляционный анализ, фильтрация. Трансформация данных, приведение их к удобному для анализа виду: метод скользящего окна, сортировка и группировка, слияние данных.	Лекционные занятия	1	2	УК-1 ПК-5
2.2	Тема 2.1. "Алгоритмы предварительной обработки данных". Подготовки данных для анализа. анализ данных. Применение факторного и корреляционного анализа.	Практические занятия	1	2	УК-1 ПК-5
2.3	Тема 2.2. "Методы Data Mining". Применение автокорреляции, регрессии, дерева решений, ассоциативных правил при анализе данных. Понятие и основные принципы «извлечения знаний из данных». Основные задачи, решаемые методами Data Mining. Автокорреляция. Линейная и логистическая регрессия. Дерево решений: принципы построения, использование, интерпретация результатов. Метод поиска ассоциативных правил, области его применения. "Нейросетевые методы анализа данных". Принципы функционирования искусственных нейронных сетей. Основные области их применения. Многослойные нейросети. Аппроксимация функциональных зависимостей. Прогнозирование временных рядов. Карты Кохонена. Задачи кластеризации и визуализации больших объемов данных. Интерпретация результатов.	Самостоятельная работа	1	10	УК-1 ПК-5
2.4	Тема "Методы Data Mining". Применение автокорреляции, регрессии, дерева решений, ассоциативных правил для анализа данных.	Самостоятельная работа	1	2	УК-1 ПК-5
2.5	Тема 2.3. "Нейросетевые методы анализа данных". Обсуждение принципов функционирования искусственных нейронных сетей. Выявление основных областей их применения.	Самостоятельная работа	1	2	УК-1 ПК-5
2.6	Тема 2.1. "Основы анализа данных". Применение метода регрессии для анализа данных. Прогнозирование и визуализация данных.	Самостоятельная работа	1	2	УК-1 ПК-5
2.7	Тема 2.4. "Инструменты многомерного статистического анализа". Статистический анализ данных. Основные возможности, спектр решаемых задач. Табличные процессоры как инструмент анализа. Эволюция инструментов: от LibreOffice Calc до SPSS и Statistica. Этапы выполнения многомерного статистического анализа и интерпретация результатов	Самостоятельная работа	1	4	УК-1 ПК-5
2.8	Подготовка реферата (темы докладов представлены в Приложении 1 РП). При подготовке реферата используется LibreOffice, MySQL Community Server.. Решение задач.	Самостоятельная работа	1	20	УК-1 ПК-5
2.9	Подготовка к промежуточной аттестации	Экзамен	1	9	УК-1 ПК-5

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	Щербаков А.	Интернет-аналитика: поиск и оценка информации в web-ресурсах: практическое пособие	Москва: Книжный мир, 2012	ЭБС «Университетская библиотека онлайн»

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
2	Пальмов, С. В.	Интеллектуальный анализ данных: учебное пособие	Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	ЭБС «IPR SMART»
3		Финансовая аналитика : проблемы и решения: журнал	Москва: Финансы и кредит, 2020	ЭБС «Университетская библиотека онлайн»
4	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

ИСС «КонсультантПлюс»

ИСС «Гарант» <http://www.internet.garant.ru/>

База статистических данных Росстата <http://www.gks.ru/>

База данных Всемирного Банка <https://datacatalog.worldbank.org/>

5.3. Перечень программного обеспечения

Операционная система РЕД ОС

LibreOffice

MySQL Community Server

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.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

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

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
УК-1: Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий			
З: правила построения баз данных, а также методики сбора данных, их проверки и фильтрации	Поиск, скачивание и анализ достоверной информации о динамике финансовых и товарных рынков с использованием современных информационных технологий с сайтов бирж и аналитических компаний при подготовке докладов и при написании рефератов	Полнота и адекватность подбора данных для выполнения поставленной исследовательской цели	Вопросы к экзамену (1-50), примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)
У: выбирать модели (алгоритмы) анализа данных; осуществлять подбор параметров модели и алгоритма обучения	Выбор наиболее адекватного метода анализа скаченной информации в зависимости от качественного состояния исследуемого рынка	Аргументированный выбор метода исследования, применяемый отечественными и зарубежными учеными;	Вопросы к экзамену (1-50), примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)
В: навыками анализа большого	Статистические параметры построенных прогнозных моделей	Критические значения статистик, с помощью которых проверяется	Вопросы к экзамену (1-50),

объёма информации, выявления скрытых закономерностей, а также прогнозирования динамики временных рядов		статистическая значимость параметров построенных моделей	примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)
ПК-5: Способен выявлять и проводить исследование актуальных научных проблем в области финансового мониторинга и финансовых рынков			
З: основные методы прикладной статистики, необходимых для проведения интеллектуального анализа данных на финансовом рынке	Демонстрирует знание основных методов прикладной и математической статистики, необходимых для проведения интеллектуального анализа данных Отвечает на вопросы, раскрывающие особенности применения конкретных статистических приемов и методов в целях осуществления Data Mining	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	Вопросы к экзамену (1-50), примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)
У: осуществлять анализ результатов проведенного исследования на основе методов прикладной статистики	Осуществляет корректное применение методов прикладной и математической статистики в процессе реализации основных этапов интеллектуального анализа данных; использует современные информационно-коммуникационные технологии и глобальные информационные ресурсы	Полнота и содержательность выполненной лабораторной работы; умение отстаивать свою позицию; обоснованность обращения к базам данных; целенаправленность поиска и отбора информации; использование различных методов и приемов прикладной и математической статистики в процессе осуществления	Вопросы к экзамену (1-50), примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)

		интеллектуального анализа данных.	
В: навыками освоения основных методов прикладной статистики в процессе осуществления Data Mining	Производит обработку, анализ, оценивание и интерпретацию результатов проведенного интеллектуального анализа данных на основе использования основных приемов и методов прикладной и математической статистики; использует современные пакеты прикладных программ для проведения расчетов; составляет отчет по лабораторной работе	Полнота и содержательность выполненной лабораторной работы, глубина анализа; использование системного подхода в процессе осуществления интеллектуального анализа данных; умение пользоваться дополнительной литературой, в том числе на иностранном языке при подготовке к занятиям; соответствие отчета по лабораторной работе предъявляемым требованиям.	Вопросы к экзамену (1-50), примерные задачи к экзамену (1-2), тесты (1-20), дискуссия (1-28), задачи (1-3), доклады (1-10), рефераты (1-32), кейс-задача (1-12)

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»)

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

Вопросы к экзамену

1. Понятие Data science.
2. Методы и стадии Data science. Свойства методов Data science.
3. Задачи, решаемые Data science. Информация и знания.
4. Очистка данных.
5. Процесс кластеризации данных. Оценка качества кластеризации.
6. Прогнозирование и кластеризация в Data science.
7. Методы и алгоритмы Data science.
8. Анализ данных в Питоне.
9. Применение корреляционного анализа для получения дополнительной информации из баз данных.
10. Применение регрессионного анализа для получения дополнительной информации из баз данных.

11. Методы классификации и прогнозирования: деревья решений.
12. Методы классификации и прогнозирования: метод опорных векторов.
13. Методы классификации и прогнозирования: метод «ближайшего соседа».
14. Методы классификации и прогнозирования: байесовская классификация.
15. Методы классификации и прогнозирования: нейронные сети.
16. Нейронные сети. Самоорганизующиеся карты Кохонена.
17. Способы визуального представления данных.
18. Data Mining консалтинг.
19. Сравнение методов сбора и обработки информации в статистике, в машинном обучении и в Data science.
20. Набор данных и их атрибутов. Типы наборов данных.
21. Базы данных. Основные положения.
22. Системы управления данными, СУБД.
23. Эволюция методов обработки и анализа данных. Переход от статистической обработки к методам OLAP и Data science.
24. Принципы, лежащие в основе методов OLAP и Data science.
25. Общая характеристика методов интеллектуального анализа данных.
26. Обзор основных задач, решаемых методами Data science.
27. Алгоритмы предварительной обработки данных.
28. Очистка данных: факторный анализ, корреляционный анализ, фильтрация.
29. Трансформация данных: метод скользящего окна, сортировка и группировка, слияние данных.
30. Понятие и основные принципы «извлечения знаний». Основные задачи, решаемые методами Data science.
31. Применение в Data science автокорреляции.
32. Применение в Data science линейной и логистической регрессии.
33. Нейросетевые методы анализа данных.
34. Принципы функционирования искусственных нейронных сетей. Основные области их применения.
35. Многослойные нейросети.
36. Аппроксимация функциональных зависимостей. Прогнозирование временных рядов.
37. Карты Кохонена. Задачи кластеризации и визуализации больших объемов данных. Интерпретация результатов.
38. Инструменты многомерного статистического анализа.
39. Статистический анализ данных. Основные возможности, спектр решаемых задач.
40. Табличные процессоры как инструмент анализа. Эволюция инструментов: от Excel до Statistica.
41. Этапы выполнения многомерного статистического анализа и интерпретация результатов.
42. Понятие базы данных, классификация данных.
43. Модели данных: иерархическая модель.
44. Модели данных: сетевая модель.
45. Модели данных: реляционная модель.
46. Структурированный язык запросов SQL: особенности, преимущества и недостатки.
47. Функции и структура СУБД.
48. Операторы создания базы данных и ввода значений.
49. Создание индекса по полям базы данных.
50. Выборка данных при помощи запросов SQL.

Примерные задачи к экзамену:

1) В файле «Задача_к_билету_1.XLS» содержатся данные с биржи ММВБ по ценам и объёму продаж обыкновенных акций Сбербанка. Каждая строка в приведенной базе данных содержит информацию об операциях за день. Задание:

- А) создайте сводную таблицу для цен и объёмов продаж ежемесячно.
- Б) постройте гистограмму для ежемесячных данных.

2) Основываясь на данных о продажах из файла «Продажи» и других сопутствующих справочниках сформировать сценарии ежедневных отчетов по долевого объему продаж в каждом из отделов по группам товаров за последние 10 дней. В каждом из отчетов должен присутствовать полный перечень товарных групп, упорядоченных по возрастанию кода.

Критерии оценивания:

- 84-100 баллов (оценка «отлично») - наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний, умений и навыков на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой;
- 67-83 балла (оценка «хорошо») - наличие определенных знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные действия по применению полученных знаний, умений и навыков на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной литературы;
- 50-66 баллов (оценка «удовлетворительно») - наличие некоторых знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; неуверенные действия по применению полученных знаний, умений и навыков на практике, грамотное и логически стройное изложение материала при ответе, усвоение с частью основной литературы;
- 0-49 баллов (оценка «неудовлетворительно») - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания, умения и навыки на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

Тесты

1. Data Mining — это процесс обнаружения в сырых данных:

- А) ранее сформулированных гипотез;
- Б) неочевидных закономерностей;
- В) практических закономерностей;
- Г) объективных закономерностей
- Д) большого количества закономерностей

2. Подготовка данных в процессе Data Mining является:

- А) необязательным этапом работы;
- Б) существенным этапом работы;
- В) может вообще отсутствовать

3. Data Mining — это процесс обнаружения в сырых данных знаний, необходимых для:

- А) принятия решений в различных сферах человеческой деятельности;
- Б) замены аналитика в процессе принятия решений;
- В) увеличения стоимости анализа данных

4. Если сравнивать Data Mining, машинное обучение и статистику, какая из дисциплин сконцентрирована на едином процессе анализа данных, включает очистку данных, обучение, интеграцию и визуализацию результатов:
- А) Data Mining;
 - Б) машинное обучение;
 - В) статистика.
5. Инструменты Data Mining:
- А) могут самостоятельно строить гипотезы о взаимосвязях в данных;
 - Б) не могут самостоятельно строить гипотезы о взаимосвязях в данных;
 - В) могут самостоятельно строить гипотезы о взаимосвязях в данных, которые обязательно подтверждаются.
6. Закономерности, найденные в процессе использования технологии Data Mining должны обладать такими свойствами:
- А) быть очевидными;
 - Б) быть неочевидными;
 - В) быть практически полезными;
 - Г) быть объективными;
 - Д) чем больше найдено закономерностей, тем лучше
7. Оцените правильность утверждения: "Data Mining может заменить аналитика"
- А) утверждение верно;
 - Б) утверждение неверно. Технология не может дать ответы на те вопросы, которые не были заданы;
 - В) утверждение неверно. Технология всего лишь дает аналитику инструмент для облегчения и улучшения его работы.
8. Какая из перечисленных ниже групп методов достаточно часто использует для выявления взаимосвязей в данных концепцию усреднения по выборке?
- А) Data Mining;
 - Б) статистические методы;
 - В) OLAP.
9. В методе опорных векторов для классификации используется ...
- А) не все множество образцов, а лишь их небольшая часть, которая находится на границах;
 - Б) все множество образцов;
 - В) часть образцов, которая находится вне границ.
10. Решаются ли задачи классификации и регрессии при помощи метода "ближайшего соседа"?
- А) да;
 - Б) нет, только задачи классификации;
 - В) нет, только задачи регрессии.
11. Метод "ближайшего соседа":
- А) может создавать модели и правила;
 - Б) может создавать модели;
 - В) может создавать правила;
 - Г) не может создавать модели и правила;

12. Назовите свойства наивной байесовской классификации:
- А) использование всех переменных и определение всех зависимостей между ними;
 - Б) наличие предположения относительно того, что все переменные являются одинаково важными;
 - В) наличие предположения относительно того, что все переменные являются статистически независимыми, т.е. значение одной переменной ничего не говорит о значении другой.
13. В синхронных нейронных сетях в каждый момент времени свое состояние меняет...
- А) лишь один нейрон;
 - Б) целая группа нейронов, как правило, весь слой;
 - В) возможен и тот и другой вариант.
14. Многослойный персептрон – это сеть ...
- А) прямого распространения сигнала (без обратных связей);
 - Б) обратного распространения сигнала (с обратными связями);
 - В) в которой входной сигнал преобразуется в выходной, проходя последовательно через несколько слоев.
15. Слой нейронной сети – это ...
- А) один или несколько нейронов, на входы которых подается один и тот же общий сигнал;
 - Б) выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов;
 - В) однонаправленные входные связи, соединенные с выходами других нейронов.
16. Сеть Кохонена представляет собой ...
- А) два слоя: входной и выходной;
 - Б) три слоя: входной, скрытый и выходной;
 - В) входной, выходной и какое угодно количество скрытых слоев.
17. Назовите характеристики кластерного анализа:
- А) не требует априорных предположений о наборе данных;
 - Б) требует априорных предположений о наборе данных;
 - В) не накладывает ограничения на представление исследуемых объектов;
 - Г) накладывает ограничения на представление исследуемых объектов;
 - Д) позволяет анализировать показатели различных типов данных.
18. Нейронные сети относятся к группам ...
- А) статистических методов;
 - Б) методов на основе уравнений;
 - В) методов кросс-табуляции.
19. Регрессионный и дискриминантный анализ ...
- А) относятся к статистическим методам Data mining;
 - Б) относятся к кибернетическим методам Data mining;
 - В) не являются методами Data mining.
20. Большинство методов Data mining были разработаны в рамках ...
- А) теории искусственного интеллекта;
 - Б) классического анализа данных;

В) теории баз данных.

Инструкция по выполнению: Тестовые задания выполняются индивидуально.

3. Критерии оценивания:

Для каждого тестового задания:

0.5 балла – дан верный ответ на тестовое задание;

0 баллов – дан неверный ответ на тестовое задание.

Максимальное количество баллов – 10.

Вопросы для дискуссии

1. Современные тенденции в области сбора, обработки и хранения больших массивов данных.
2. Методы формализации знаний и извлечения знаний у экспертов.
3. Возможности и проблемы построения интеллектуального пространства корпорации.
4. Методы сбора и обработки данных в консалтинговой фирме.
5. Структура и функции системы управления данными в коммерческом банке.
6. Эволюция методов анализа данных: исторический обзор
7. Методы построения моделей сложных систем.
8. Модель черного ящика.
9. Основные этапы моделирования.
10. Методика анализа данных.
11. Определения OLAP, Data science и взаимосвязи между ними.
12. Особенности OLAP.
13. Аналитическая отчетность и многомерное представление данных.
14. Хранилище данных.
15. Измерения и факты.
16. Основные операции над кубом данных.
17. Типы задач, решаемые методами Data science.
18. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.
19. Этапы подготовки данных.
20. Выдвижение гипотез.
21. Методы сбора и систематизации фактов.
22. Методы проведения экспертиз для выявления наиболее значимых факторов.
23. Понятия парциальной и комплексной обработки.
24. Анализ качества полученных моделей.
25. Критерии выбора аналитических платформ и пакетов Data science.
26. Основные этапы внедрения систем анализа данных.
27. Категории пользователей аналитических систем.
28. Способы снижения рисков проектов Data science.

Инструкция по выполнению: дискуссии выполняются коллективно.

3. Критерии оценивания:

Для каждого вопроса:

1 балл – дан верный ответ на вопрос;

0 баллов – дан неверный ответ.
Максимальное количество баллов – 10.

Задачи

1. Задача на поиск ассоциативных правил.

С сайта <https://biconsult.ru/products/algorithm-apriori-s-ispolzovaniem-python> скачать базу данных Groceries_dataset.csv о проданных товарах некоторой фирмой. Ознакомиться на этом же сайте с алгоритмом майнинга ассоциаций на основе данных транзакций из розничной торговой площадки или интернет-магазина. Проанализировать корзину продаж с помощью алгоритма Априори с использованием Python, Построить график десяти самых продаваемых товаров.

2. Задача по кластеризации данных.

Для решения задач кластеризации данных использовать Python, а также библиотеку scikit-learn для загрузки и обработки набора данных и matplotlib для визуализации. Базу данных (набор данных ирисов Фишера) для выполнения кластеризации скачать с сайта <https://proglab.io/p/unsupervised-ml-with-python>. На этом же сайте ознакомиться с алгоритмом кластеризации. Построить три кластера для трех видов цветов: Setosa, Versicolor и Virginica.

3. Задача на построение прогнозного регрессионного уравнения с использованием программы Python.

Импортировать пакет NumPy и класс LinearRegression из библиотеки sklearn. Для шести пар точек (x, y) построить линейное регрессионное уравнение вида:

$$Y = a + b \cdot x + e$$

С исходным программным кодом можно ознакомиться на сайте <https://proglab.io/p/linear-regression>

Рассчитать в Python коэффициенты «a» и «b», а также коэффициент детерминации.

Инструкция по выполнению: задачи выполняются индивидуально на компьютере.

Инструкция по выполнению: задачи выполняются индивидуально.

Критерии оценивания:

Для каждой задачи:

10 баллов – задача решена правильно;

7-9 баллов – задача решена с небольшими неточностями;

4-6 баллов – неправильно интерпретированы полученные результаты;

1-3 балла – решение задачи содержит существенные ошибки;

0 баллов – задача решена неправильно.

Максимальное количество баллов – 30.

Примерные темы докладов

1. Алгоритмы Data Mining: дерево решений.
2. Алгоритмы Data Mining: кластеризация.
3. Алгоритмы Data Mining: наивный алгоритм Байеса.
4. Алгоритмы Data Mining: временные ряды.

5. Алгоритмы Data Mining: ассоциативные правила.
6. Алгоритмы Data Mining: кластеризация последовательностей действий.
7. Алгоритмы Data Mining: нейронные сети.
8. Алгоритмы Data Mining: линейная регрессия.
9. Алгоритмы Data Mining: логистическая регрессия.
10. Получение информации при помощи SQL-запросов.

Критерии оценивания:

Каждый доклад оценивается максимум в 5 баллов:

- 4-5 баллов - системность, обстоятельность и глубина излагаемого материала; знакомство с научной и учебной литературой; способность воспроизвести основные тезисы доклада без помощи конспекта; способность быстро и развернуто отвечать на вопросы преподавателя и аудитории; наличие презентации к докладу;
- 3 балла - развернутость и глубина излагаемого в докладе материала; знакомство с основной научной литературой к докладу; при выступлении частое обращение к тексту доклада; некоторые затруднения при ответе на вопросы (неспособность ответить на ряд вопросов из аудитории); наличие презентации;
- 1-2 балла - правильность основных положений доклада; наличие недостатка информации в докладе по целому ряду проблем; использование для подготовки доклада исключительно учебной литературы; неспособность ответить на несложные вопросы из аудитории и преподавателя; неумение воспроизвести основные положения доклада без письменного конспекта; наличие презентации;
- 0 баллов - поверхностный, неупорядоченный, бессистемный характер информации в докладе; при чтении доклада постоянное использование текста; полное отсутствие внимания к докладу аудитории; отсутствие презентации.

Максимальное количество баллов – 10 (за 2 доклада).

Тематика рефератов

1. Информация как источник данных».
2. Анализ данных, знаний и структур в системах искусственного интеллекта
3. Искусственный интеллект в современном мире»
4. Визуализация как средство анализа информации
5. Применение ассоциативных правил в бизнесе».
6. Перспективные средства анализа бизнес-информации
7. Методы анализа структурных объектов
8. Современные IT-решения в визуализации данных
9. Анализ текстовой информации»
10. Сравнительная характеристика методов анализа данных
11. Требования, предъявляемые к OLAP-системам
12. Задачи и содержание оперативного (OLAP) анализа
13. Сравнительный анализ моделей информационных систем OLTP и OLAP
14. Витрины данных. Варианты использования
15. Область применения BI.
16. Для решения каких практических задач используется прогнозирование?
17. Классификация. Основные понятия и определения. Правила и деревья классификации.
18. Многомерный статистический анализ данных
19. Методы и области применения интеллектуального анализа Data mining. ПО Data mining.

20. Data Mining. Основные понятия и определения. Шкалы. Задачи анализа данных.
21. Алгоритмы Data Mining Дерево решений - Microsoft Decision Trees
22. Алгоритмы Data Mining Кластеризация - Microsoft Clustering
23. Алгоритмы Data Mining Наивный алгоритм Байеса - Microsoft Native Bayes
24. Алгоритмы Data Mining Временные ряды - Microsoft Time Series
25. Алгоритмы Data Mining Ассоциативные правила - Microsoft Association
26. Алгоритмы Data Mining Кластеризация последовательностей действий
Microsoft Sequence Clustering
27. Алгоритмы Data Mining Нейронные сети - Microsoft Neural Network
28. Алгоритмы Data Mining Линейная регрессия - Microsoft Linear Regression
29. Алгоритмы Data Mining Логистическая регрессия - Microsoft Logistic Regression
30. Средства преобразования данных
31. Средства оперативного (OLAP) анализа
32. Средства интеллектуального анализа данных

Критерии оценивания:

- 20-30 баллов - содержание работы полностью соответствует теме; глубоко и аргументировано раскрывается тема; логическое и последовательное изложение мыслей; написано правильным литературным языком и стилистически соответствует содержанию; заключение содержит выводы, логично вытекающие из содержания основной части;
- 10-19 баллов - достаточно полно и убедительно раскрывается тема с незначительными отклонениями от нее; в основной части логично, связно, но недостаточно полно доказывается выдвинутый тезис; имеются незначительные нарушения последовательности в изложении мыслей; заключение содержит выводы, логично вытекающие из содержания основной части;
- 1-9 баллов - в основном раскрывается тема; дан верный, но односторонний или недостаточно полный ответ на тему; обнаруживается недостаточное умение делать выводы и обобщения; выводы не полностью соответствуют содержанию основной части;
- 0 баллов - тема полностью нераскрыта; характеризуется случайным расположением материала, отсутствием связи между частями; выводы не вытекают из основной части; многочисленные заимствования текста из других источников.

Максимальное количество баллов – 30

Кейс-задача

1. Проанализируйте влияние инструментов корпоративного управления на изменение финансово-экономических показателей деятельности компаний.
2. Проведите анализ текущего состояния российских “голубых фишек” методом мультипликаторов.
3. Проведите сравнение эффективности использования метода мультипликаторов при определении инвестиционной привлекательности акций на фондовых рынках России и США.
4. Покажите возможность применения мультипликаторов для анализа различных отраслей российского фондового рынка».
5. Проведите анализ инвестиционной привлекательности российских компаний на основе метода мультипликаторов.

6. Покажите влияние ли кросс-листинг на стоимость российских компаний?
7. Проанализируйте опережающие показатели финансовых кризисов.
8. Охарактеризуйте альтернативные инвестиции: современный рынок искусства.
9. Проанализируйте возможности выхода российских инвесторов на зарубежные финансовые рынки.
10. Определите детерминанты доходности корпоративных облигаций при размещении.
11. Покажите взаимодействие шоков ликвидности и волатильности на фондовом рынке для акций компаний различной капитализации.
12. Покажите слияние инвестиций в интеллектуальный капитал на результаты деятельности компаний.

Критерии оценивания:

Каждое задание оценивается максимум в 5 баллов:

- 4-5 баллов - системность, обстоятельность и глубина излагаемого материала; знакомство с научной и учебной литературой; способность воспроизвести основные тезисы доклада без помощи конспекта; способность быстро и развернуто отвечать на вопросы преподавателя и аудитории; наличие презентации к докладу;
- 3 балла - развернутость и глубина излагаемого в докладе материала; знакомство с основной научной литературой к докладу; при выступлении частое обращение к тексту доклада; некоторые затруднения при ответе на вопросы (неспособность ответить на ряд вопросов из аудитории); наличие презентации;
- 1-2 балла - правильность основных положений доклада; наличие недостатка информации в докладе по целому ряду проблем; использование для подготовки доклада исключительно учебной литературы; неспособность ответить на несложные вопросы из аудитории и преподавателя; неумение воспроизвести основные положения доклада без письменного конспекта; наличие презентации;
- 0 баллов - поверхностный, неупорядоченный, бессистемный характер информации в докладе; при чтении доклада постоянное использование текста; полное отсутствие внимания к докладу аудитории; отсутствие презентации.

Максимальное количество баллов – 10 (за 2 выполненные задачи).

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 экзамена.

Экзамен проводится по расписанию промежуточной аттестации в устном виде. Количество вопросов в задании – 3. 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.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- лабораторные занятия.

В ходе лекционных занятий рассматриваются вопросы, относящиеся к поиску, скачиванию, анализу, статистической обработке, хранению и визуализации больших объёмов данных.

В ходе практических занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки сбора, анализа и синтеза информации.

При подготовке к лабораторным занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;
- изучить конспекты лекций;
- подготовить ответы на все вопросы по изучаемой теме.

По согласованию с преподавателем студент может подготовить реферат по теме занятия и презентацию к нему, используя LibreOffice. В процессе подготовки к лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и практических занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий посредством тестирования. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников. Выделить непонятные термины, найти их значение в энциклопедических словарях.

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 рефератов, требования к оформлению

Общие положения

Реферат это одна из форм устной итоговой аттестации. Реферат – это самостоятельная исследовательская работа, в которой автор раскрывает суть исследуемой проблемы; приводит различные точки зрения, а также собственные взгляды на нее. Содержание реферата должно быть логичным; изложение материала носит проблемно-тематический характер.

Требования к написанию реферата

Автор реферата должен продемонстрировать достижение им уровня мировоззренческой, общекультурной компетенции, т.е. продемонстрировать знания по выбранному вопросу данного курса.

1. Необходимо правильно сформулировать тему, отобрать по ней необходимый материал.
2. Использовать только тот материал, который отражает сущность темы.
3. Во введении к реферату необходимо обосновать выбор темы.
4. После цитаты необходимо делать ссылку на автора.
5. Изложение должно быть последовательным. Недопустимы нечеткие формулировки, речевые и орфографические ошибки.
6. В подготовке реферата необходимо использовать материалы современных изданий не старше 5 лет.
7. Оформление реферата (в том числе титульный лист, литература) должно быть грамотным.
8. Список литературы оформляется с указанием автора, названия источника, места издания, года издания, названия издательства, использованных страниц.

Требования к оформлению реферата

- Изложение текста и оформление реферата, список литературы выполняют в соответствии с требованиями ГОСТ Р 7.0.5 2008. Страницы текстовой части и включенные в нее иллюстрации и таблицы должны соответствовать формату А4 по ГОСТ 9327-60.
- Реферат должен быть выполнен печатным способом на одной стороне бумаги формата А4 через полтора интервала. Цвет шрифта должен быть черным, высота букв, цифр и других знаков не менее 1.8 (шрифт 14 пт.).
- Текст следует печатать, соблюдая следующие размеры полей: верхнее и нижнее — 20 мм, левое — 30 мм, правое — 10 мм. Абзацный отступ должен быть одинаковым по всему тексту и составлять 1,25 см.
- Выравнивание текста по ширине.
- Разрешается использовать компьютерные возможности акцентирования внимания на определенных терминах, формулах, применяя выделение жирным шрифтом, курсив, подчеркивание.
- Перенос слов недопустим!
- Точку в конце заголовка не ставят. Если заголовок состоит из двух предложений, их разделяют точкой.
- Подчеркивать заголовки не допускается.
- Название каждой главы и параграфа в тексте работы можно писать более крупным шрифтом, жирным шрифтом, чем весь остальной текст. Каждая глава начинается с новой страницы, параграфы (подразделы) располагаются друг за другом.
- В тексте реферат рекомендуется чаще применять красную строку, выделяя законченную мысль в самостоятельный абзац.

- Все страницы обязательно должны быть пронумерованы. Нумерация листов должна быть сквозной. Номер листа проставляется арабскими цифрами.
- Нумерация листов начинается с третьего листа (после содержания) и заканчивается последним. На третьем листе ставится номер «3».
- Номер страницы на титульном листе не проставляется!
- Номера страниц проставляются в центре нижней части листа без точки. Список использованной литературы и приложения включаются в общую нумерацию листов.
- Рисунки и таблицы, расположенные на отдельных листах, включают в общую нумерацию листов и помещают по возможности следом за листами, на которых приведены ссылки на эти таблицы или иллюстрации. Таблицы и иллюстрации нумеруются последовательно арабскими цифрами сквозной нумерацией. Допускается нумеровать рисунки и таблицы в пределах раздела. В этом случае номер таблицы (рисунка) состоит из номера раздела и порядкового номера таблицы, разделенных точкой.

Оформление литературы:

Каждый источник должен содержать следующие обязательные реквизиты:

- фамилия и инициалы автора;
- наименование;
- издательство;
- место издания;
- год издания.

Все источники, включенные в библиографию, должны быть последовательно пронумерованы и расположены в следующем порядке:

- законодательные акты (кодексы, федеральные законы, указы Президента;
- постановления Правительства;
- нормативные документы;
- статистические материалы;
- научные и литературные источники – в алфавитном порядке по первой букве фамилии автора/названия книги под редакцией.

В конце работы размещаются приложения. В тексте на все приложения должны быть даны ссылки. Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова «Приложение» и его номера. Приложение должно иметь заголовок, который записывают симметрично относительно текста с прописной буквы отдельной строкой.

Содержание реферата

1. Титульный лист должен содержать наименование учреждения, в котором выполнялся реферат, Ф.И.О. автора, Ф.И.О. руководителя, год написания.

2. В реферате в обязательном порядке должно быть представлено:

Введение (не более 2-3 стр.), где отражены:

- цели и задачи работы
- основной замысел.

Основная часть (15-20 стр.), обусловлена задачами исследования

Заключительные выводы (1,5-2 стр.)

Список литературы

Приложения (таблицы, графики, схемы).