

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Директор

Дата подписания: 30.01.2024 17:26:33

Уникальный программный ключ:

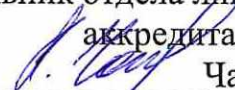
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Начальник отдела лицензирования и аккредитации



Чаленко К.Н.

« 30 » августа 2021 г.

**Рабочая программа дисциплины
Введение в Data Mining**

по профессионально-образовательной программе направление 01.03.05 "Статистика"
профиль 01.03.05.01 "Анализ больших данных"

Для набора 2021 года


Квалификация
Бакалавр


КАФЕДРА **Статистики, эконометрики и оценки рисков**

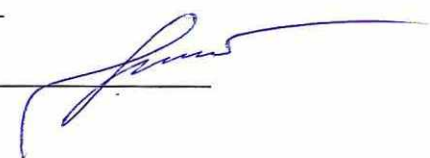
Семестр (<Курс>.<Семестр на курсе>)	2 (1.2)		Итого	
	16			
Неделя	16			
Вид занятий	уп	рп	уп	рп
Лекции	16	16	16	16
Лабораторные	16	16	16	16
Итого ауд.	32	32	32	32
Контактная работа	32	32	32	32
Сам. работа	76	76	76	76
Часы на контроль	36	36	36	36
Итого	144	144	144	144

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 30.08.2021 протокол № 1.

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

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

Методическим советом направления: к.э.н., доц., Кислая И.А. 

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

ПК-4:Способен формировать источники больших данных

В результате освоения дисциплины обучающийся должен:

Знать:
Правила построения баз данных, а также методики сбора данных, их проверки и фильтрации;
Уметь:
выбирать модели (алгоритмы) анализа данных; осуществлять подбор параметров модели и алгоритма обучения;
Владеть:
анализа большого объема информации, выявления скрытых закономерностей, а также прогнозирования динамики временных рядов.

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	Раздел 1. Data mining и базы данных				
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. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
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. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
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. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
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. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3

1.5	Тема 1.3. "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 распределенности. Сверхбольшие базы данных. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
1.6	Тема 1.3. "Правила построения баз данных". Построение баз данных в СУБД Microsoft SQL Server. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
1.7	Тема 1.4. "Введение в SQL-запросы". Простейшие Select- запросы. Получение необходимой информации при помощи SQL -запросов. Формирование баз данных. Манипулирование данными: использование подзапросов Insert, Delete и Update. Использование индексации для быстрого доступа к данным. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
1.8	Тема 1.4. "Введение в SQL-запросы". Составление простейших Select-запросов. Формирование базы данных с заданными характеристиками и требуемой информацией. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
1.9	Разработка связанных баз данных и набора SQL-запросов, которые бы помогали собирать информацию с конкретного сегмента финансового рынка для выработки инвестиционных решений. /Ср/	2	36	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
Раздел 2. Методы классификации и прогнозирования данных					
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. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.2	Тема 2.1. "Алгоритмы предварительной обработки данных". Подготовки данных для анализа. анализ данных. Применение факторного и корреляционного анализа. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3

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. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.4	Тема "Методы Data Mining". Применение автокорреляции, регрессии, дерева решений, ассоциативных правил для анализа данных. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.5	Тема 2.3. "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. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.6	Тема 2.3. "Нейросетевые методы анализа данных". Обсуждение принципов функционирования искусственных нейронных сетей. Выявление основных областей их применения. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.7	Тема 2.1. "Основы анализа данных". Классификация. Регрессия. Методы представления результатов анализа. Прогнозирование и визуализация. Методы прогнозирования и классификации: алгоритм построения 1R-алгоритм; метод Naïve Bayes; деревья решений, метод опорных векторов, метод «ближайшего соседа», нейронные сети. Поиск ассоциативных правил. Характеристики ассоциативных правил: поддержка, достоверность и улучшение. Алгоритм Apriori. Методы кластерного анализа. Меры близости в алгоритмах кластеризации. /Лек/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.8	Тема 2.1. "Основы анализа данных". Применение метода регрессии для анализа данных. Прогнозирование и визуализация данных. /Лаб/	2	2	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.9	Тема 2.4. "Инструменты многомерного статистического анализа". Статистический анализ данных. Основные возможности, спектр решаемых задач. Табличные процессоры как инструмент анализа. Эволюция инструментов: от Excel до SPSS и Statistica. Этапы выполнения многомерного статистического анализа и интерпретация результатов /Ср/	2	20	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
2.10	Подготовка реферата (темы докладов представлены в Приложении 1 РП). При подготовке реферата используется Microsoft Office, Excel, MS SQL. Решение задач. /Ср/	2	20	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3

2.11	Экзамен /Экзамен/	2	36	ПК-4	Л1.6 Л1.1 Л1.7 Л1.2 Л1.3 Л1.4 Л1.8 Л1.5Л2.6 Л2.1 Л2.7 Л2.5 Л2.2 Л2.4 Л2.3
------	-------------------	---	----	------	---

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

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

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1		Data Mining	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016	http://www.iprbookshop.ru/56315.html неограниченный доступ для зарегистрированных пользователей
Л1.2	Уэс Маккинли, Слинкин А. А.	Python и анализ данных	Саратов: Профобразование, 2017	http://www.iprbookshop.ru/64058.html неограниченный доступ для зарегистрированных пользователей
Л1.3	Пальмов, С. В.	Интеллектуальный анализ данных: учебное пособие	Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017	http://www.iprbookshop.ru/75376.html неограниченный доступ для зарегистрированных пользователей
Л1.4	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	http://www.iprbookshop.ru/81324.html неограниченный доступ для зарегистрированных пользователей
Л1.5		Финансовая аналитика : проблемы и решения: журнал	Москва: Финансы и кредит, 2020	https://biblioclub.ru/index.php?page=book&id=574735 неограниченный доступ для зарегистрированных пользователей
Л1.6	Щербаков А.	Интернет-аналитика: поиск и оценка информации в web-ресурсах: практическое пособие	Москва: Книжный мир, 2012	https://biblioclub.ru/index.php?page=book&id=89693 неограниченный доступ для зарегистрированных пользователей
Л1.7	Нестеров, С. А.	Интеллектуальный анализ данных средствами MS SQL Server 2008	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016	http://www.iprbookshop.ru/62813.html неограниченный доступ для зарегистрированных пользователей
Л1.8	Мельниченко, А. С.	Математическая статистика и анализ данных: учебное пособие	Москва: Издательский Дом МИСиС, 2018	http://www.iprbookshop.ru/78563.html неограниченный доступ для зарегистрированных пользователей

5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1		Журнал "Вопросы статистики"	,	1
Л2.2	Пол Дейтел, Харви Дейтел	Python: Искусственный интеллект, большие данные и облачные вычисления	Санкт-Петербург: Питер, 2021	https://ibooks.ru/reading.php?short=1&productid=371701 неограниченный доступ для зарегистрированных пользователей
Л2.3	Барсегян А., Куприянов М., Степаненко В., Холод И.	Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. 2 изд.	Санкт-Петербург: БХВ-Петербург, 2008	https://ibooks.ru/reading.php?short=1&productid=335156 неограниченный доступ для зарегистрированных пользователей
Л2.4	Барсегян А., Куприянов М., Степаненко В. Холод И.	Методы и модели анализа данных: OLAP и Data Mining	Санкт-Петербург: БХВ-Петербург, 2004	https://ibooks.ru/reading.php?short=1&productid=335137 неограниченный доступ для зарегистрированных пользователей
Л2.5	Рассел Мэтью, Классен Михаил	Data Mining. Извлечение информации из Facebook, Twitter, LinkedIn, Instagram. GitHub. 3-е изд.	Санкт-Петербург: Питер, 2020	https://ibooks.ru/reading.php?short=1&productid=365266 неограниченный доступ для зарегистрированных пользователей
Л2.6		Журнал "Финансовые исследования"	,	1
Л2.7	Адлер, Ю. П., Черных, Е. А.	Статистическое управление процессами. «Большие данные»: учебное пособие	Москва: Издательский Дом МИСиС, 2016	http://www.iprbookshop.ru/64199.html неограниченный доступ для зарегистрированных пользователей

5.3 Профессиональные базы данных и информационные справочные системы

КонсультантПлюс

Гарант

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

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

5.4. Перечень программного обеспечения

Microsoft Office

Excel

MS SQL

5.5. Учебно-методические материалы для студентов 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 (МОДУЛЯ)

Помещения для проведения всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения. Для проведения лекционных занятий используется демонстрационное оборудование.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

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			
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	В – вопросы экз. (1-50), Т - тесты (1-20); ДС – дискуссии (темы 1-28); Д – доклады (1-10); Р – рефераты (темы 1-32)
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;	В – вопросы экз. (1-50), 3 – задачи (1-3); Кейс-задачи (1-12)
В: анализа большого объема информации, выявления скрытых закономерностей, а также прогнозирования	Статистические параметры построенных прогнозных моделей	Критические значения статистик, с помощью которых проверяется статистическая значимость параметров построенных моделей	В – вопросы экз. (1-50), 3 – задачи (1-3); Кейс-задачи (1-12)

динамики временных рядов.			
---------------------------	--	--	--

3.2 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 по дисциплине: «Введение в Data Mining»

1. Понятие Data Mining.
2. Методы и стадии Data Mining. Свойства методов Data Mining.
3. Задачи, решаемые Data Mining. Информация и знания.
4. Очистка данных.
5. Процесс кластеризации данных. Оценка качества кластеризации.
6. Прогнозирование и кластеризация в Data Mining.
7. Методы и алгоритмы Data Mining.
8. Анализ данных в MS Excel.
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 Mining.
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 Mining.
24. Принципы, лежащие в основе методов OLAP и Data Mining.
25. Общая характеристика методов интеллектуального анализа данных.

26. Обзор основных задач, решаемых методами Data Mining.
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 Mining.
31. Применение в Data Mining автокорреляции.
32. Применение в Data Mining 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.

Пример теста
по дисциплине: «Введение в Data Mining»

1. Какие из нижеперечисленных признаков являются количественными:
 - а) плотность населения
 - б) уровень занятости населения
 - в) среднедушевой доход
 - г) пол человека
 - д) возраст
 - е) уровень образования (начальное, среднее, высшее)
2. Какие из нижеперечисленных признаков являются дискретными:
 - а) денежные доходы населения
 - б) число детей в семье
 - в) прибыль предприятия
 - г) пол человека
 - д) тарифный разряд
3. Типологические группировки применяются для:
 - а) характеристики структурных сдвигов
 - б) характеристики взаимосвязей между отдельными признаками
 - в) разделения совокупности на качественно однородные типы
 - г) характеристики структуры совокупности
4. Структурные группировки применяются для:
 - а) разделения совокупности на качественно однородные типы
 - б) характеристики взаимосвязей между отдельными признаками
 - в) характеристики структуры совокупности
5. Часть зависимой переменной в регрессионной модели, которая может быть объяснена значением регрессора:
 - а) случайное возмущение;
 - б) отклик;
 - в) уравнение регрессии;
 - г) остаток.

6. Гипотеза является сложной, если:
- она состоит из конечного числа простых гипотез;
 - она состоит из бесконечного числа простых гипотез;
 - Содержит только одно предположение.
7. Коррелированность возмущений с различными номерами называется:
- гомоскедастичностью;
 - гетероскедастичностью;
 - автокорреляцией.
8. Критической областью называют:
- совокупность значений критерия, при которых нулевую гипотезу принимают;
 - совокупность значений критерия, при которых нулевую гипотезу отвергают.
9. Причины гетероскедастичности (множественный выбор):
- исследование неоднородных объектов;
 - характер наблюдений;
 - ошибки спецификации;
 - ошибки измерений.
10. Под мультиколлинеарностью понимается линейная зависимость (единичный выбор):
- зависимой переменной с одним или несколькими регрессорами;
 - двух, или нескольких регрессоров;
 - зависимой переменной с возмущением;
 - регрессоров с возмущением.
11. С увеличением объема выборки длина доверительного интервала прогнозируемого значения зависимой переменной (единичный выбор):
- увеличивается;
 - уменьшается;
 - не меняется.
12. Как переводится DATA MINING?
- “добыча” или “раскопка данных”.
 - “значение”.
 - “хранение”.
 - “перечисление данных”.
13. Какое требование к переработке информации не верно?
- Данные имеют неограниченный объем.
 - Данные являются разнородными.
 - Результаты должны быть конкретны и понятны.
 - Инструменты для обработки сырых данных должны быть сложны в использовании.
14. Какая концепция положена в основу современной технологии Data Mining?
- Концепция естествознания.
 - Концепция управления.
 - Концепция шаблонов (паттернов).
 - Концепция становления.
15. Сколько выделяют стандартных типов закономерностей?
- 4.

- 5.
 - 6.
 - 7.
16. Если несколько событий связаны друг с другом, то это...
- Ассоциация.
 - Последовательность.
 - Классификация.
 - Кластеризация.
17. Основой для каких систем служит историческая информация, хранящаяся в БД в виде временных рядов?
- Классификации.
 - Последовательности.
 - Прогнозирования.
 - Ассоциации.
18. Какую иерархическую структуру создают деревья решений?
- "ЕСЛИ... ТО...".
 - "НИ... НИ...".
 - "КОГДА... ТО...".
 - "... НИКОГДА...".
19. С чем связано направление эволюционного программирования?
- Постановка вопроса вида "значение параметра А больше х?".
 - Использование метода "ближайшего соседа".
 - Подача значений входных параметров, на основе которых нужно принимать какие-то решения, прогнозировать развитие ситуации.
 - Поиск зависимости целевых переменных от остальных в форме функций какого-то определенного вида.
20. Что называют хромосомами в генетических алгоритмах?
- Кодировку исходных логических закономерностей в базе данных.
 - Направление эволюционного программирования.
 - Большой класс систем.
 - Набор закономерностей.

Инструкция по выполнению: Тестовые задания выполняются индивидуально.

3. Критерии оценки:

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
по дисциплине: «Введение в Data Mining»

Студент выбирает 1-2 вопроса из приведенного списка:

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 Mining 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 Mining.
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 Mining.
26. Основные этапы внедрения систем анализа данных.
27. Категории пользователей аналитических систем.
28. Способы снижения рисков проектов Data Mining.

Инструкция по выполнению: дискуссии выполняются коллективно.

3. Критерии оценки:

Для каждого вопроса:

1 балл – дан верный ответ на вопрос;

0 баллов – дан неверный ответ.

Максимальное количество баллов – 10.

Задачи

7

1. Коммерческий директор хочет иметь информацию о последних тенденциях в изменении суммовых объемов продаж по товарным группам. Для этого предлагается вычислить относительное изменение объемов продаж за последние 10 дней по отношению к объемам продаж за предыдущие 10 дней. Если это изменение менее -0,3, то товарной группе присваиваем категорию «Провал», от -0,3 до -0,1 – «Падение», от -0,1 до 0 – «Уменьшение», от 0 до 0,1 – «Увеличение», от 0,1 до 0,3 – «Подъем», свыше 0,3 – «Взлет». Необходимые данные находятся в файлах «Продажи», «Товары», «Товарные группы».
2. Вы работаете в небольшой туристической фирме и планируете массовую рассылку рекламного буклета. Ваши средства ограничены, поэтому вы хотите послать ее тем, кто готов тратить на путешествия и отдых в большей степени. В файле КЛИЕНТЫ.XLS содержатся данные о случайной выборке клиентов размером 30 (пол, возраст, суммы, затраченные на путешествия и отдых в предыдущем году). Используйте данные, чтобы понять, насколько пол и возраст влияют на объем затрат. Сформулируйте обоснованные рекомендации относительно контингента для рассылки рекламной брошюры.
3. Основываясь на данных о продажах из файла «Продажи» и других сопутствующих справочниках сформировать сценарий ежедневного отчета по 5 лидерам товарных групп по суммам продаж за последние 15 дней. Отчет визуализировать с помощью столбчатых диаграмм с информацией о сумме продаж и названии товарных групп.

Инструкция по выполнению: задачи выполняются индивидуально.

Критерии оценки:

Для каждой задачи:

10 баллов – задача решена правильно;

0 баллов – задача решена неправильно.

Максимальное количество баллов – 30.

Примерные темы докладов

по дисциплине: «Введение в Data Mining»

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 оценки:

Каждый доклад оценивается максимум в 5 баллов:

- 4-5 баллов - системность, обстоятельность и глубина излагаемого материала;

8

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 по дисциплине: «Введение в Data Mining»

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 оценки:

- 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 не раскрыта; 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 по дисциплине: «Введение в Data Mining»

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 оценки:

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 экзаменационном задании – 3. Объявление результатов производится в день экзамена. Результаты аттестации заносятся в экзаменационную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

Приложение 2

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 объемов данных.

В ходе лабораторных 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 Microsoft Office. В процессе подготовки к лабораторным занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и лабораторных занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом устного опроса или 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 выполняются 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 Times New Roman, 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).