

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Узунов Федор Владимирович
Должность: Ректор
Дата подписания: 04.02.2022 13:10:13
Уникальный программный ключ:
fd935d10451b860e912264c0378f8448452bfdb603f9a3a57460185

**АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
«ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ»
«УНИВЕРСИТЕТ ЭКОНОМИКИ И УПРАВЛЕНИЯ»
Факультет экономики и управления
Кафедра «Бизнес-информатика»**



УТВЕРЖДАЮ

Проректор по учебно-методической работе

С.С. Скараник

«01» сентября 2021 г.

Рабочая программа дисциплины

Анализ больших данных

Направление подготовки
38.03.05 Бизнес-информатика

Квалификация выпускника
Бакалавр

Для всех
форм обучения

Симферополь 2021

АННОТАЦИЯ	
Индекс дисциплины по учебному плану	Наименование дисциплины
Б1.О.30	Анализ больших данных
Цель изучения дисциплины	Формирование у обучающихся представления об основных методах и программных средствах обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений, навыков работы с большими объемами информации.
Место дисциплины в структуре ОПОП	Дисциплина относится к обязательной части блока 1. «Дисциплины (модули)» программы бакалавриата
Компетенции, формируемые в результате освоения дисциплины	ОПК-4
Содержание дисциплины	Тема 1. Интеллектуальный анализ данных Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных Тема 3. Статистическое мышление и статистический анализ Тема 4. Методы классификации и прогнозирования Тема 5. Введение в машинное обучение и разработку данных. Тема 6. Применение технологий больших данных для бизнес-анализа Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации
Общая трудоемкость дисциплины	Общая трудоемкость дисциплины составляет 4 зачетных единиц (144 часов)
Форма промежуточной аттестации	Зачет с оценкой

Содержание

1. Цель и перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы бакалавриата	5
2. Место дисциплины в структуре ОПОП бакалавриата	5
3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся	5
4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	6
5. Контроль качества освоения дисциплины	10
6. Учебно-методическое обеспечение дисциплины	10
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	10
8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	11
9. Методические указания для обучающихся по освоению дисциплины	11
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)	12
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	12
Приложение 1. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	13

1. Цель и перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы бакалавриата

Целью изучения дисциплины «Анализ больших данных» является формирование у обучающихся представления об основных методах и программных средствах обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений, навыков работы с большими объемами информации.

В результате освоения ОПОП бакалавриата обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенции	Результаты освоения ОПОП	Перечень планируемых результатов обучения по дисциплине
ОПК-4	Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений	ОПК-4.1 Знает методы и программные средства сбора, обработки и анализа информации; ОПК-4.2 Умеет работать с большими объемами информации, использовать методы и программные средства обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений; ОПК-4.3 Владеет программными средствами сбора, обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений.

2. Место дисциплины в структуре ОПОП бакалавриата

Дисциплина Б1.О.30 «Анализ больших данных» относится к обязательной части блока 1. «Дисциплины (модули)» учебного плана ОПОП бакалавриата по направлению подготовки 38.03.05 Бизнес-информатика. Дисциплина «Анализ больших данных» изучается обучающимися очной формы обучения в 6 семестре, очно-заочной формы обучения – в 7 семестре.

При изучении данной дисциплины обучающийся использует знания, умения и навыки, которые сформированы в процессе изучения предшествующих дисциплин: «Эконометрика», «Специальные разделы математики», «Экономико-математическое моделирование», «Информационные системы и технологии в экономике и управлении», «Моделирование бизнес-процессов», «Нечеткая логика и нейронные сети».

Знания, умения и навыки, полученные при изучении дисциплины «Анализ больших данных», будут необходимы для углубленного и осмысленного восприятия дисциплин: «Информационная безопасность и защита информации», «Системы поддержки и методы принятия решений», «Основы цифровой экономики».

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Общая трудоемкость (объем) дисциплины составляет 4 зачетных единицы (з.е.), 144 академических часов.

3.1. Объем дисциплины по видам учебных занятий (в часах)

Для очной формы обучения

Общая трудоемкость дисциплины составляет 4 зачётных единицы 144 часов

Объём дисциплины	Всего часов
Общая трудоемкость дисциплины	144
Контактная работа	44
Аудиторная работа (всего):	44
Лекции	22
Семинары, практические занятия	22
Самостоятельная работа обучающихся (всего)	100
Курсовая работа	-
Зачет с оценкой	+
Экзамен	-

Для очно-заочной формы обучения

Общая трудоемкость дисциплины составляет 4 зачётных единицы 144 часов

Объём дисциплины	Всего часов
Общая трудоемкость дисциплины	144
Контактная работа	34
Аудиторная работа (всего):	34
Лекции	16
Семинары, практические занятия	18
Самостоятельная работа обучающихся (всего)	110
Курсовая работа	-
Зачет с оценкой	+
Экзамен	-

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины и трудоемкость по видам учебных занятий (в академических часах)

№ темы	Наименование темы	Всего		Количество часов	
		Ф	З	Контактная работа	Внеаудит. работа

				Лекции		Практические		Самост. работа	
				ОФО	ОЗФО	ОФО	ОЗФО	ОФО	ОЗФО
1.	Интеллектуальный анализ данных	24	24	4	4	4	4	16	16
2.	Сбор, хранение, обработка, извлечение и анализ больших потоков данных	24	24	4	2	4	2	16	20
3.	Статистическое мышление и статистический анализ	20	20	2	2	2	2	16	16
4.	Методы классификации и прогнозирования	24	24	4	2	4	2	16	20
5.	Введение в машинное обучение и разработку данных	24	24	4	2	4	4	16	18
6.	Применение технологий больших данных для бизнес-анализа	14	14	2	2	2	2	10	10
7.	Способы визуализации, интерпретации, проектная организация и бизнес рекомендации	14	14	2	2	2	2	10	10
	Всего по дисциплине	144	144	22	16	22	18	100	110
	Контроль								
	Итого	144	144	22	16	22	18	100	110

4.2. Содержание дисциплины, структурированное по темам (разделам)

Тема 1. Интеллектуальный анализ данных

Цели и задачи дисциплины. Введение в Data mining. Определение больших данных. Методы и стадии Data Mining. Задачи Data Mining. Извлечение, данных, информации и знаний. Типы задач, решаемые методами Data Mining: классификация, кластеризация, регрессия, ассоциация, поиск последовательных шаблонов. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.

Сферы применения Data Mining. Научные проблемы в области больших данных. Современные программные средства анализа больших данных.

Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных

Основные характеристики и типы больших данных. Технологии хранения больших данных. Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

Технологии обработки больших данных. Технологии анализа больших данных.

Технологии Business Intelligence и реляционные системы управления базами данных.

Тема 3. Статистическое мышление и статистический анализ

Статистическое мышление. Обработка данных для контроля качества процессов и технологии sixsigma. Статистический подход к анализу данных. Точечное и интервальное оценивание. Методы статистического анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).

Специализированные статистические пакеты.

Тема 4. Методы классификации и прогнозирования

Анализ последовательностей. Отбор признаков и снижение размерности. Работа с выбросами и пропущенными значениями. Классификация и кластеризация. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы.

Прогнозирование и визуализация данных.

Тема 5. Введение в машинное обучение и разработку данных.

Основные аналитические методы обработки данных. Машинное обучение и майнинг больших данных (Big Data).

Нейронные сети как реализация алгоритмов машинного обучения. Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура.

Тема 6. Применение технологий больших данных для бизнес-анализа

Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях, в масс медиа, FinTech как область применения бизнес анализа.

Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов Data Mining.

Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации

Способы визуального представления данных. Методы визуализации. Базовые принципы визуализации. Инструменты визуализации данных.

Визуальное управление проектами.

Основные тенденции в области визуализации

Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

4.3.Содержание практических занятий (очная форма обучения)

Практическое занятие 1-2. Интеллектуальный анализ данных (4 часа)

1. Основные понятия и определения Data Mining.
2. Методы и стадии Data Mining. Задачи Data Mining. Типы задач, решаемые методами Data Mining
3. Сферы применения Data Mining. Современные программные средства анализа больших данных.

Практическое занятие 3-4. Сбор, хранение, обработка, извлечение и анализ больших потоков данных (4 часа)

1. Обзор технологий хранения больших данных.

<p>2. Модели данных.</p> <p>3. Технологии обработки больших данных. Технологии анализа больших данных. Технологии Business Intelligence и реляционные системы управления базами данных.</p>
<p>Практическое занятие 4. Статистический анализ (2 часа)</p> <p>1. Методы статистического анализа данных</p> <p>2. Специализированные статистические пакеты</p>
<p>Практическое занятие 5-6. Методы классификации и прогнозирования (4 часа)</p> <p>1. Анализ последовательностей.</p> <p>2. Классификация и кластеризация. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы.</p> <p>3. Прогнозирование и визуализация данных.</p>
<p>Практическое занятие 7-8. Введение в машинное обучение и разработку данных (4 часа)</p> <p>1. Машинное обучение и майнинг больших данных (Big Data).</p> <p>2. Нейронные сети как реализация алгоритмов машинного обучения.</p> <p>3. Многослойный перцептрон: класс решаемых задач, архитектура.</p> <p>4. Построение нейросетевой прогнозной модели спроса.</p>
<p>Практическое занятие 9. Применение технологий больших данных для бизнес-анализа (2 часа)</p> <p>1. Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях, в масс медиа, FinTech как область применения бизнес анализа.</p> <p>2. Построение скоринговой модели кредитования (дерева решений);</p>
<p>Практическое занятие 10. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации. (2 часа)</p> <p>1. Способы визуального представления данных. Базовые принципы визуализации. Инструменты визуализации данных. Методы визуализации.</p> <p>2. Визуальное управление проектами.</p> <p>3. Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.</p>

4.4. Содержание самостоятельной работы

<p>Тема 1. Интеллектуальный анализ данных</p> <p>1. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.</p> <p>2. Сферы применения Data Mining. Научные проблемы в области больших данных.</p>
<p>Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных</p> <p>1. Технологии обработки больших данных. Технологии анализа больших данных.</p> <p>2. Технологии Business Intelligence и реляционные системы управления базами данных.</p>
<p>Тема 3. Статистическое мышление и статистический анализ</p> <p>1. Методы статистического анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).</p>

<p>Тема 4. Методы классификации и прогнозирования</p> <ol style="list-style-type: none"> 1. Мультимодальная кластеризация. Предиктивные алгоритмы. 2. Прогнозирование и визуализация данных.
<p>Тема 5. Введение в машинное обучение и разработку данных</p> <ol style="list-style-type: none"> 1. Нейронные сети как реализация алгоритмов машинного обучения. Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура.
<p>Тема 6. Применение технологий больших данных для бизнес-анализа</p> <ol style="list-style-type: none"> 1. Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов Data Mining.
<p>Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации</p> <ol style="list-style-type: none"> 1. Основные тенденции в области визуализации 2. Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

5. Контроль качества освоения дисциплины

Текущий контроль и промежуточная аттестация осуществляются в соответствии с «Положением о текущей и промежуточной аттестации обучающихся в Автономной некоммерческой организации «Образовательная организация высшего образования» «Университет экономики и управления».

Вид промежуточной аттестации – зачет с оценкой.

Оценочные средства по дисциплине приведены в Приложении 1.

6. Учебно-методическое обеспечение дисциплины

1. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/89482.html>. — Режим доступа: для авторизир. пользователей

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная литература:

1. Чубукова, И. А. Data Mining : учебное пособие / И. А. Чубукова. — 3-е изд. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 469 с. — ISBN 978-5-4497-0289-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/89404.html>. — Режим доступа: для авторизир. пользователей
2. Пальмов, С. В. Интеллектуальный анализ данных : учебное пособие / С. В. Пальмов. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/75376.html>. — Режим доступа: для авторизир. пользователей

3. Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ-МГСУ, ЭБС АСВ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/101802.html>. — Режим доступа: для авторизир. пользователей
- б) дополнительная литература:**
4. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/88752.html>. — Режим доступа: для авторизир. пользователей
5. Синева, И. С. Анализ данных в среде R. Ч. 1 : учебное пособие / И. С. Синева. — Москва : Московский технический университет связи и информатики, 2018. — 32 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/92422.html>. — Режим доступа: для авторизир. пользователей
6. Брусенцев, А. Г. Анализ данных и процессов. Ч.1. Методы статистического анализа данных : учебное пособие / А. Г. Брусенцев. — Белгород : Белгородский государственный технологический университет им. В.Г. Шухова, ЭБС АСВ, 2017. — 63 с. — ISBN 978-5-361-00540-6. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/92237.html>. — Режим доступа: для авторизир. пользователей
7. Афанасьева, Т. В. Введение в проектирование систем интеллектуального анализа данных : учебное пособие / Т. В. Афанасьева, А. Н. Афанасьев. — Ульяновск : Ульяновский государственный технический университет, 2017. — 64 с. — ISBN 978-5-9795-1686-8. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/106086.html>. — Режим доступа: для авторизир. пользователей

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. <http://www.iprbookshop.ru/>- Электронно-библиотечная система IPRbooks
2. <http://www.window.edu.ru> - Единое окно доступа к образовательным ресурсам. Полнотекстовая электронная библиотека учебных и учебно-методических материалов (федеральный ресурс)
3. <http://www.olap.ru> – информационный портал, посвященный технологиям интерактивной аналитической обработки
4. <https://proglib.io/p/neural-nets-guide> – информационный портал «Библиотека программиста»

9. Методические указания для обучающихся по освоению дисциплины

При проведении лекций, семинарских (практических) занятий, самостоятельной работе обучающихся применяются интерактивные формы проведения занятий с целью погружения обучающихся в реальную атмосферу профессионального сотрудничества по разрешению проблем, оптимальной выработки навыков и качеств будущего специалиста. Интерактивные формы проведения занятий предполагают обучение в сотрудничестве. Все

участники образовательного процесса (преподаватель и обучающиеся) взаимодействуют друг с другом, обмениваются информацией, совместно решают проблемы, моделируют ситуацию.

В учебном процессе используются интерактивные формы занятий:

- творческое задание. Выполнение творческих заданий требует от обучающегося воспроизведение полученной ранее информации в форме, определяемой преподавателем, и требующей творческого подхода;

- групповое обсуждение. Групповое обсуждение кого-либо вопроса направлено на достижение лучшего взаимопонимания и способствует лучшему усвоению изучаемого материала.

В ходе освоения дисциплины при проведении контактных занятий используются следующие формы обучения, способствующие формированию компетенций: лекции-дискуссии; кейс-метод; решение задач; ситуационный анализ; обсуждение рефератов и докладов; разработка групповых проектов; встречи с представителями государственных и общественных организаций.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

В процессе лекционных и практических занятий используется следующее программное обеспечение:

*программы, обеспечивающие доступ к «Интернет» (например, «Googlechrome»);

*программы, демонстрации видео материалов (например, проигрыватель «Windows MediaPlayer»);

*программы для демонстрации и создания презентаций (например, «MicrosoftPowerPoint»).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для преподавания дисциплины не требуется специальных материально-технических средств (лабораторного оборудования, компьютерных классов и т.п.). Однако во время лекционных занятий, которые проводятся в большой аудитории, использовать проектор для демонстрации слайдов, схем, таблиц и прочего материала, мультимедийные проекторы Epson, BenqViewSonic; экраны для проекторов; ноутбуки Asus, Lenovo, микрофоны.

**Фонд оценочных средств для проведения промежуточной аттестации
обучающихся по дисциплине**

1. Типовые контрольные задания или иные материалы

1.1 Примерные вопросы к зачету с оценкой

1. Введение в Data mining. Определение больших данных.
2. Методы и стадии Data Mining.
3. Задачи Data Mining. Извлечение, данных, информации и знаний.
4. Типы задач, решаемые методами Data Mining: классификация, кластеризация, регрессия, ассоциация, поиск последовательных шаблонов.
5. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.
6. Сферы применения Data Mining. Научные проблемы в области больших данных.
7. Современные программные средства анализа больших данных.
8. Основные характеристики и типы больших данных.
9. Технологии хранения больших данных. Обзор технологий хранения больших данных.
10. Модели данных.
11. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.
12. Технологии обработки больших данных.
13. Технологии анализа больших данных.
14. Технологии Business Intelligence и реляционные системы управления базами данных.
15. Обработка данных для контроля качества процессов и технологии sixsigma.
16. Статистический подход к анализу данных. Точечное и интервальное оценивание.
17. Методы статистического анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).
18. Специализированные статистические пакеты.
19. Анализ последовательностей. Отбор признаков и снижение размерности. Работа с выбросами и пропущенными значениями.
20. Классификация и кластеризация. Поиск признаковых зависимостей и частых множеств.
21. Мультимодальная кластеризация. Предиктивные алгоритмы.
22. Основные аналитические методы обработки данных. Машинное обучение и майнинг больших данных (Big Data).
23. Нейронные сети как реализация алгоритмов машинного обучения. Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей.
24. Многослойный перцептрон: класс решаемых задач, архитектура.
25. Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях, в масс медиа, FinTech как область применения бизнес анализа.

26. Способы визуального представления данных. Методы визуализации.
27. Базовые принципы визуализации. Инструменты визуализации данных.
28. Визуальное управление проектами.
29. Основные тенденции в области визуализации
30. Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

1.2 Типовые темы рефератов

1. Алгоритмы кластеризации.
2. Нечеткие методы представления данных.
3. Роль аналитика по данным (Data Scientist).
4. Программное обеспечение SAS Enterprise Miner.
5. Распознавание образов.
6. Роль больших данных в бизнесе.
7. Нейронные сети как реализация алгоритмов машинного обучения.
8. Применение технологий больших данных для задач управления в реальном времени.
9. Модели распределенных файловых систем. Файловая система Google и Hadoop.
10. MapReduce. Парадигма, суть, структура.
11. Поточная модель данных Алгоритм Флажолет-Мартина. Алгоритм Алона-Матиаса Жегеди. Алгоритм Датара-Гиониса-Индика-Мотвани (DGIM).
12. Анализ связей. Page Rank.
13. Задачи классификации, как инструмент повышения эффективности бизнеса.
14. Методология построения и верификации прогнозных моделей.
15. Методы оценки качества прогнозных моделей.

1.3. Пример типовых тестовых заданий

1. Аналитик это ...

- а) специалист в области анализа и моделирование
- б) специалист в предметной области;
- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

2 Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

3 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.

4 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.

5 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.

6 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.

7 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.

8 К описательным моделям 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.

9 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.

10. Модели последовательностей 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.

11 Регрессивные модели 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 анализу.

12. Модели ассоциации проявляют ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

13 Виды физической неопределенности данных:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью)
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);
- г) неоднозначность смысла фраз (Синтаксическая и семантическая).

14 Очистка данных — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

15 Консолидация — ...

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

16 Классификация — ...

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

17 Регрессия — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

18 Кластеризация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных.

19 Ассоциация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

20 Машинное обучение — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной влияние, что и отвечает ему правильный выходной результат.

г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

21 Подготовка данных в процессе Data Mining является:

а) необязательным этапом работы.

б) существенным этапом работы.

в) может вообще отсутствовать

22 Какая концепция положена в основу современной технологии Data Mining?

а) Концепция естествознания.

б) Концепция управления.

в) Концепция шаблонов (паттернов).

г) Концепция становления.

23 Если несколько событий связаны друг с другом, то это...

- а) Ассоциация.
- б) Последовательность.
- в) Классификация.
- г) Кластеризация

24 Основой для каких систем служит историческая информация, хранящаяся в БД в виде временных рядов?

- а) Классификации.
- б) Последовательности.
- в) Прогнозирования.
- г) Ассоциации.

25 Какую иерархическую структуру создают деревья решений?

- а) "ЕСЛИ... ТО...".
- б) "НИ... НИ...".
- в) "КОГДА... ТО...".
- г) "... НИКОГДА...".

26 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

27 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

28 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

29 Решаются ли задачи классификации и регрессии при помощи метода "ближайшего соседа"?

- а) да
- б) нет, только задачи классификации
- в) нет, только задачи регрессии

30 Выберите характеристику, наиболее подходящую для Data Mining

- а) подходит для понимания ретроспективных данных
- б) опирается на ретроспективные данные для получения ответов на вопросы о будущем
- в) подходит для обобщения ретроспективных данных

