

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт экономики и менеджмента

УТВЕРЖДАЮ:
Директор Е. В. Нехода

Рабочая программа дисциплины

Анализ данных

по направлению подготовки

38.03.02 Менеджмент

Направленность (профиль) подготовки:
Менеджмент

Форма обучения
Очная

Квалификация
Бакалавр

Год приема
2020

СОГЛАСОВАНО:
Руководитель ОП И.А. Павлова
Председатель УМК В.В. Макоеева

Томск – 2020

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен осуществлять сбор, обработку и анализ данных, необходимых для решения поставленных управленческих задач, с использованием современного инструментария и интеллектуальных информационно-аналитических систем;

ПК-4 Способен осуществлять деятельность по обеспечению персоналом.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.4 Использует современный инструментарий и интеллектуальные информационно-аналитические системы для решения поставленных управленческих задач

ИПК-4.1 Анализирует рынок труда и формирует требования к вакантной должности

ИПК-4.2 Использует разнообразные методы поиска, привлечения, подбора и отбора персонала в соответствии с утвержденными планами на основе анализа затрат

2. Задачи освоения дисциплины

– Изучить особенности хранения и компьютерной обработки данных и основные подходы к их анализу.

– Научиться применять понятийный аппарат теории баз данных и интеллектуального анализа данных для эффективного решения практических задач профессиональной деятельности.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплина (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения.

Код дисциплины в учебном плане: Б1.В.03

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Пятый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Информационные технологии», основные дисциплины математического блока.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 4 з.е., 144 часов, из которых:

-лекции: 24 ч.

-практические занятия: 30 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в анализ данных

Анализ данных. Модель. Моделирование. Свойства модели. Аналитический подход к моделированию. Информационный подход к моделированию.

Тема 2. Процесс анализа

Извлечение и визуализация данных. Построение моделей. Тиражирование знаний. Подготовка данных к анализу.

Тема 3. Технологии KDD и DM

Базовые классы задач. Бизнес-задачи, которые решаются алгоритмами DataMining. Аналитические платформы.

Тема 4. Основные понятия баз данных

Понятие базы данных. Понятие СУБД. Типовая организация СУБД. Понятие модели данных. Основные модели данных: иерархическая, сетевая, реляционная, объектная. Трехуровневая архитектура СУБД, трехуровневое представление данных. Логическая и физическая независимость данных.

Тема 5. Проектирование реляционных баз данных

Цели проектирования. Понятие избыточных данных. Нормальные формы: свойства нормальных форм. Нормализация отношений. I, II, III нормальные формы.

Тема 6. Язык SQL

Основные характеристики языка. Диалекты языка. Группы операторов SQL.

Тема 7. Архитектуры СУБД и информационных систем

Работа СУБД в вычислительных сетях. Архитектура построения сетевых СУБД. Архитектура «файл-сервер» и «клиент/сервер». Двухуровневая и трехуровневая архитектура систем баз данных. Распределенная обработка данных. Модель удаленного доступа к данным. Модель сервера баз данных. Модель сервера приложений. Виды схем построение ИС распределенной обработки информации.

Тема 8. Консолидация данных

Основные задачи консолидации данных. Процесс консолидации. Хранилища данных. Транзакция. Требования к ХД. Основные положения концепции ХД. Многомерные хранилища данных. Реляционные хранилища данных. Гибридные хранилища данных. Витрины данных. ETL.

Тема 9. Трансформация данных

Трансформация данных на разных этапах аналитического процесса. Основные методы трансформации данных.

Тема 10. Алгоритмы DataMining

Машинное обучение. Обучение с учителем и без учителя.

Обучающее и тестовое множества. Ошибка обобщения. Эффект переобучения. Вычислительная сложность алгоритмов. Понятие масштабируемых алгоритмов

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, решения задач, тестов по изученному материалу, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр. Вклад результатов текущего контроля в итоговой оценке по дисциплине составляет – 60 баллов (60%).

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в пятом семестре проводится в письменной форме по билетам. Экзаменационный билет состоит из двух частей. Продолжительность экзамена 1,5 часа. Ответ на вопрос первой части дается в развернутой форме.

Вторая часть содержит вопросы, оформленные в виде практической задачи. Ответы на вопросы второй части предполагают решение задачи и краткую интерпретацию полученных результатов.

Результаты экзамена определяются в соответствии с балльно-рейтинговой системой – максимум 40 баллов за экзамен (40%):

Критерии выставления баллов за письменный экзамен:

Баллы	Характеристика
40 баллов	Даны полные и развернутые ответы на вопросы. Задача решена верно и дана обоснованная интерпретация полученных результатов.
30 баллов	Даны неполные ответы на вопросы. Задача решена верно, но интерпретация полученных результатов не убедительна.
20 баллов	Даны фрагментарные ответы на вопросы. Задача решена верно, но интерпретация полученных результатов не убедительна.
10 баллов	Даны ограниченные ответы на вопросы. Задача решена неверно, но была попытка интерпретации полученных результатов.

Итоговая оценка по дисциплине складывается из результатов текущего контроля (60%) и результатов промежуточной аттестации (40%) и составляет максимум 100 баллов.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Механизм перевода результатов балльно-рейтинговой системы в пятибалльную шкалу:

Баллы	Итоговая оценка
85-100 баллов	«Отлично»
70-84 балла	«Хорошо»
55-69 баллов	«Удовлетворительно»
54 балла и менее	«Неудовлетворительно»

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=33711>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

Примерный перечень теоретических вопросов:

1. Понятия «Модель» и «Моделирование». Основные подходы к моделированию.
2. Процесс анализа данных.
3. Технология KDD. Место, роль и основные задачи DM.
4. Назначение и основные компоненты системы БД: понятия «информация» и «данные»; понятие банка данных (БД); понятие «предметная область»; понятия базы данных БД, словаря данных, базы знаний.

5. Обзор современных систем управления базами данных (СУБД). Функция СУБД. Трёхуровневая структура СУБД.
6. Концептуальная схема БД. Требования, предъявляемые к концептуальной схеме БД. Правила построения концептуальной схемы БД. Модели представления концептуальных схем БД
7. Этапы проектирования базы данных: цели, задачи, особенности и результаты каждого из этапов.
8. Обеспечение целостности данных. Проблемы целостности данных. Типы ограничений целостности.
9. Классификация моделей данных. Достоинства и недостатки моделей.
10. Понятие модели данных, типы структур данных, операции над данными, ограничения целостности.
11. Реляционная модель данных: схема отношения. Язык манипулирования данными; ключи; связывание таблиц.
12. Реляционная модель данных: основные понятия, ограничения целостности, особенности манипулирования данными, достоинства и недостатки. Примеры СУБД, основанных на реляционной модели данных.
13. Основные этапы проектирования БД. Основные проблемы, возникающие в процессе проектирования. Понятие «модели данных». Схема данных.
14. Хранилища данных. Требования к ХД. Основные положения концепции ХД. Многомерные хранилища данных. Реляционные хранилища данных. Гибридные хранилища данных. Витрины данных.
15. Процесс ETL.
16. Трансформация данных. Основные методы трансформации данных.
17. Алгоритмы DataMining

Примерный перечень практических задач, проверяющих индикаторы:

Задача №1. На основе таблицы Data создать таблицу AutoColorTop12, содержащую все поля исходной таблицы, в которую отбираются записи об автомобилях, окрашенных в один из 12 наиболее популярных цветов. В полученной таблице AutoColorTop12 снизить цену на 15% всем подержанным отечественным автомобилем, выпущенным до 2000 года.

Задача №2. На основе таблицы Data создать перекрестный запрос с вычислением среднего среднегодового пробега по каждой модели 5-й и 7-й серии (модель начинается с цифры 5 или 7) автомобилей марки BMW с годом выпуска с 1995 по 2001 (включительно) из городов Москва и Владимир, в котором строки - Модель, столбцы - ГодВыпуска, значения - Средний Пробег.

Задача №3. Проведите аудит данных, ответив на следующие вопросы:

1. Есть ли в данных дубликаты, противоречия, пропуски, аномалии?
2. Какова доля неполных и некорректных записей в общем объеме?
3. Какие поля представляют интерес для анализа?
4. Интегральная оценка качества данных.

Задача №4. В базе данных education_process.mdb хранятся данные о результатах работы студентов. Спроектируйте и создайте хранилище данных. Создайте запрос на выборку информации о количестве и проценте решенных задач каждого упражнения по Информатике каждым студентом.

в) Методические указания по организации самостоятельной работы студентов:

Самостоятельная работа включает в себя:

– самостоятельную подготовку к занятиям по заявленным темам курса в соответствии с содержанием дисциплины и литературой. Контроль выполнения производится на занятиях в блиц-опросах;

- самостоятельную работу в аудитории при ответах на вопросы, решении задач. Контроль выполнения осуществляется сразу же при оценке полученных результатов;
- самостоятельное выполнение индивидуальных аналитических заданий. Контроль выполнения осуществляется в сроки, предусмотренные для сдачи индивидуальных заданий, которые оговариваются со студентами;
- самостоятельную подготовку к экзамену. Контроль выполнения заключается в проставлении итоговой оценки по итогам обучения.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Авдеенко С.Н. Управление данными средствами MS Office. Электронный ресурс: Учебно-методический комплекс /С. Н. Авдеенко, А. Л. Богданов. - Томск: ИДО ТГУ, 2010. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000405022>
- Базы данных : учебник для прикладного бакалавриата / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. – 3-е изд., перераб. и доп. – Москва : Издательство Юрайт, 2019. – 420 с.
- Замятин А. В. Интеллектуальный анализ данных : учебное пособие : [для студентов университетов и вузов] / А. В. Замятин ; Нац. исслед. Том. гос. ун-т. - Томск : Издательский Дом Томского государственного университета, 2020. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000722107>

б) дополнительная литература:

- Нестеров, С. Л. Базы данных : учебник и практикум для вузов / С. Л. Нестеров. - Москва : Издательство Юрайт, 2020. – 230 с.
- Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. - Москва: Издательство Юрайт, 2020. – 174с
- Анализ данных: учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. – Москва : Издательство Юрайт, 2020. – 490 с.
- Шнарсва, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарсва, Ж. Г. Пономарева. – Симферополь : Университет экономики и управления, 2019. – 129 с.

в) ресурсы сети Интернет:

- Электронный ресурс по базам данных <http://citforum.ru/database/>
- Электронный ресурс по базам данных <http://www.sql.ru/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office Access, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- Аналитическая платформа Deductor 5.2 (5.3);
- публично доступные облачные технологии (GoogleDocs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>

- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Авдеенко Сергей Николаевич, кандидат технических наук, доцент кафедры информационных технологий и бизнес-аналитики Института экономики и менеджмента Национального исследовательского Томского государственного университета.