

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Сахалинский государственный университет»

Кафедра информатики

УТВЕРЖДЕН
на заседании кафедры
«19» марта 2024 г., протокол № 8
Исполняющий обязанности
заведующего кафедрой



Осипов Г.С.

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ
ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

Б1.В.ДВ.03.02 Введение в аналитику данных

Направление подготовки
10.03.01 Информационная безопасность

профиль

Безопасность автоматизированных систем (по отрасли или в сфере профессиональной
деятельности)

**Уровень высшего образования
БАКАЛАВРИАТ**

Южно-Сахалинск
2024 г.

1. Формируемые компетенции и индикаторы их достижения по дисциплине

Код компетенции	Содержание компетенции	Код и наименование индикатора достижения компетенции
ПКС-1	Способен проводить формализацию предметной области с целью создания информационной системы в сфере профессиональной деятельности	ПКС-1.1 - Знает критерии оценки эффективности и надежности средств защиты программного обеспечения автоматизированных систем; ПКС-1.2 - Умеет определять параметры настройки программного обеспечения системы защиты информации автоматизированной системы; ПКС-1.3 – Владеет навыками определения параметров настройки программного обеспечения системы защиты информации автоматизированной системы;

2. Паспорт фонда оценочных средств по дисциплине (модуля)

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных.	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
2.	Визуализация данных. Представление результатов исследования.	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
3.	Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ.	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
4.	Анализ сетей	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
5.	Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
6.	Деревья решений. Случайный лес. Ансамбли моделей.	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету
7.	зачет	ПКС-1	Задания к лабораторным работам, контрольные вопросы, вопросы к зачету

Лабораторное занятие №1 (4 ч.)

Тема Тема 1. Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных.

Вопросы для обсуждения:

1. Математический аппарат для анализа данных: векторы, матрицы, функции и производные.
2. Особенные типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).
3. Модуль для работы с числовыми данными NumPy (Numerical Python). Особенные типы данных в NumPy. Работа с векторами и матрицами.
4. Модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.
5. Вычисление главных статистических метрик с помощью NumPy. Работа с табличными данными в Pandas
6. Проведение разведывательного анализа данных в Pandas на примере данных о пассажирах “Титаника” (базовый тренировочный набор данных).

Лабораторное занятие №2 (4 ч.)

Тема Визуализация данных. Представление результатов исследования

Вопросы для обсуждения:

1. Нюансы визуализации данных и принципы человеческого восприятия.
2. Правила создания хороших визуализаций.
3. Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования.
4. Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных. Создание визуализаций для разведывательного анализа данных для набора данных о пассажирах “Титаника”. Работа с данными Gapminder в Pandas (<https://www.gapminder.org>): разведывательный анализ, преобразование данных.
5. Воссоздание визуализации Gapminder (зависимость продолжительности жизни от ВВП на душу населения).
6. Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly. (создание тепловых географических карт, интерактивного пузырькового графика).

Лабораторное занятие №3 (6 ч.)

Тема Работа с текстовыми данными. Сбор данных из открытых источников. Предварительная обработка текстовых данных. Текстовый анализ

Вопросы для обсуждения:

1. Введение в анализ текста. Применение в политологии. Особенности подготовки данных. Анализ текста. Латентное размещение Дирихле.
2. Сбор данных для текстового анализа с новостных сайтов. Подготовка данных для анализа. Использование моделей Word2vector. Анализ текста, собранного из социальной сети (комментарии пользователей).
3. Построение предсказательных моделей.

Лабораторное занятие №4 (6 ч.)

Тема Анализ сетей

Вопросы для обсуждения:

1. Основные метрики и параметры сетей.
2. Модуль NetworkX.
3. Подготовка данных для анализа сетей.
4. Примеры визуализации сетей на примере данных из социальной сети.

Лабораторное занятие №5 (6 ч.)

Тема Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.

Вопросы для обсуждения:

1. Основные понятия машинного обучения.
2. Обучение с учителем и без учителя.
3. Проблема переобучения.
4. Регрессионные модели.
5. Метод наименьших квадратов.
6. Логистическая регрессия.
7. Решение задач кластеризации. Меры расстояния.
8. Алгоритмы кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связанных компонент).
9. Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”.
10. Построчный разбор алгоритма кластеризации. Проверка его работы на синтетических данных (ирисы, случайные формы).
11. Применение алгоритма кластеризации к набору данных.

Лабораторное занятие №6 (6 ч.)

Тема Деревья решений. Случайный лес. Ансамбли моделей

Вопросы для обсуждения:

1. Введение в ансамбли моделей. Стэкинг и бэггинг.
2. Случайный лес.
3. Бустинг.
4. Разбор примеров. Построение ансамблей моделей на наборе данных “Титаник”

Форма контроля – **зачет**

Задания для текущего контроля

Примеры домашних заданий

1. Импортируйте набор данных о пассажирах “Титаника”. Проведите разведывательный анализ данных. Поставьте гипотезу. Проведите необходимые преобразования данных.
2. На основании поставленной гипотезы в ДЗ1 выберите модели для предсказания. Обоснуйте выбор. Постройте выбранные модели, выберите лучшую. Постройте ансамбль моделей. Улучшились ли предсказания? Оформите результаты в виде отчета.
3. Вариант проекта: соберите данные из социальной сети или с новостного сайта для последующего текстового анализа. Проведите предварительную обработку данных. Поставьте задачу классификации. Проведите необходимые преобразования данных. Постройте модель. Оцените, насколько хорошо модель справляется с поставленной задачей. Опишите результат, назовите возможные причины плохой работы модели (в случае, если результаты неудовлетворительные).

Примерные вопросы к зачету.

1. Математический аппарат для анализа данных: векторы, матрицы, функции и производные.
2. Особенные типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).
3. Модуль для работы с числовыми данными NumPy (Numerical Python). Особенные типы данных в NumPy. Работа с векторами и матрицами.
4. Модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.
5. Правила создания хороших визуализаций.
6. Введение в анализ текста. Особенности подготовки данных. Анализ текста. Латентное размещение Дирихле.
7. Построение предсказательных моделей.
8. Основные метрики и параметры сетей.
9. Модуль NetworkX.
10. Подготовка данных для анализа сетей.
11. Примеры визуализации сетей на примере данных из социальной сети.
12. Основные понятия машинного обучения.
13. Обучение с учителем и без учителя.
14. Проблема переобучения.
15. Регрессионные модели.
16. Метод наименьших квадратов.
17. Логистическая регрессия.
18. Решение задач кластеризации. Меры расстояния.
19. Алгоритмы кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связанных компонент).
20. Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”.
21. Построчный разбор алгоритма кластеризации. Проверка его работы на синтетических данных (ирисы, случайные формы).
22. Применение алгоритма кластеризации к набору данных.
23. Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг.

3. Система оценивания планируемых результатов обучения

Критерии оценивания:

Критерием оценивания является выполнение самостоятельных заданий, контрольных и лабораторных работ.

Самостоятельные задания, контрольные и лабораторные работы по результатам выполнения и защиты оцениваются с учетом следующих основных параметров:

- своевременное выполнение работы;
- полнота и правильность ответов на вопросы, заданные в ходе защиты работы.

В случае выполнения данных условий, студент имеет возможность сдавать теоретический зачет по вопросам.

Оценка «зачтено» выставляется студенту, глубоко и прочно усвоившему программный материал, исчерпывающе, последовательно, грамотно и логически стройно его излагающему, в ответе которого увязывается теория с практикой, он показывает знакомство с литературой, правильно обосновывает и использует рациональные и современные средства решения поставленной проблемы.

оценка «**не зачтено**» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, допускающему в ответе или в решении задач грубые ошибки.

Форма контроля	За одну работу		Всего	
	Мин. баллов	Макс. баллов	Мин. баллов	Макс. баллов
Текущий контроль:				
Активная работа на занятии	0,25	0,5	9	18
Выполнение домашнего задания	0,75	0,75	27	27
Выполнение заданий самостоятельной работы	1	3	1	3
коллоквиум	1	3	3	9
Промежуточная аттестация (зачет)			20	43
Итого за семестр			60	100

Составитель Н. Вашидзе Вашакидзе Н.С., старший преподаватель кафедры информатики

«07» марта 2024 г.