

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Комсомольский-на-Амуре государственный университет»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ¹
по дисциплине

Интеллектуальный анализ данных

Направление подготовки	<i>09.04.03 «Прикладная информатика»</i>
Направленность (профиль) образовательной программы	<i>Интеллектуальные системы</i>

Обеспечивающее подразделение
<i>Кафедра ПУРИС – Проектирование, управление и разработка информационных систем</i>

Разработчик ФОС:

Доцент, кандидат технич. наук
(должность, степень, ученое звание)

(подпись)

А.В. Инзарцев
(ФИО)

Оценочные материалы по дисциплине рассмотрены и одобрены на заседании
кафедры, протокол № _____ от « ____ » _____ 2023 г.

Заведующий кафедрой _____ А.Н.Петрова

¹ В данном документе представлены типовые оценочные средства. Полный комплект оценочных средств, включающий все варианты заданий (тестов, контрольных работ и др.), предлагаемых обучающемуся, хранится на кафедре в бумажном и электронном виде.

1 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами образовательной программы

Таблица 1 – Компетенции и индикаторы их достижения

Код и наименование компетенции	Индикаторы достижения	Планируемые результаты обучения по дисциплине
Общепрофессиональные		
ОПК-3. Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями;	ОПК-3.1. Знает принципы, методы и средства анализа и структурирования профессиональной информации; ОПК-3.2. Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров; ОПК-3.3. Владеет навыками использования современного программного обеспечения для анализа данных и компьютерного моделирования.	Знает: - принципы, методы и средства анализа и структурирования профессиональной информации Умеет: - анализировать профессиональную информацию, выделять в ней главное, структурировать Владеет: - навыками использования современного программного обеспечения для анализа данных и компьютерного моделирования

Таблица 2 – Паспорт фонда оценочных средств

Контролируемые разделы (темы) дисциплины	Формируемая компетенция	Наименование оценочного средства	Показатели оценки
Основы анализа данных.	ОПК-3	Лабораторные работы	Знает принципы, методы и средства анализа и структурирования профессиональной информации Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать Владеет навыками использования современного программного обеспечения для анализа данных и компьютерного моделирования
Корреляционно-регрессионный анализ.	ОПК-3	Лабораторные работы	Знает принципы, методы и средства анализа и структурирования профессиональной информации

			Умеет анализировать профессиональную информацию с использованием корреляционно-регрессионного анализа, выделять в ней главное, структурировать.
Классификация данных	ОПК-3	Лабораторные работы	Знает принципы, методы и средства анализа и структурирования профессиональной информации Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать Владеет навыками использования современного программного обеспечения для анализа данных и компьютерного моделирования
Кластерный анализ	ОПК-3	Лабораторные работы	Знает принципы, методы и средства анализа и структурирования профессиональной информации Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать Владеет навыками использования современного программного обеспечения для анализа данных и компьютерного моделирования
	ОПК-3	РГР	Умеет анализировать профессиональную информацию с использованием корреляционно-регрессионного анализа, выделять в ней главное, структурировать.

2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующие процесс формирования компетенций

Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, представлены в виде технологической карты дисциплины (таблица 3).

Таблица 3 – Технологическая карта

	Наименование оценочного средства	Сроки выполнения	Шкала оценивания	Критерии оценивания
Промежуточная аттестация в форме Зачёта с оценкой				
	Лабораторные работы	В течение трёх недель с даты выдачи	10 баллов (за каждую из 12 лабораторных работ)	- Выполнено без ошибок и в срок – 10 баллов; - Нарушены сроки сдачи – минус 1 балл; - Допущены погрешности не принципиального характера – минус 1 балла; - Допущены незначительные ошибки, исправленные под руководством преподавателя – минус 2 балла.
	Расчётно-графическая работа	В течение пяти недель с даты выдачи	30 баллов	- Выполнено без ошибок и в срок – 30 баллов; - Нарушены сроки сдачи – минус 3 балла; - Допущены погрешности не принципиального характера – минус 4 балла; - Допущены незначительные ошибки, исправленные под руководством преподавателя – минус 7 баллов.
	ИТОГО:	-	150 баллов	-
Критерии оценки результатов обучения по дисциплине: 0 – 64 % от максимально возможной суммы баллов – «неудовлетворительно» (недостаточный уровень для промежуточной аттестации по дисциплине); 65 – 74 % от максимально возможной суммы баллов – «удовлетворительно» (пороговый (минимальный) уровень); 75 – 84 % от максимально возможной суммы баллов – «хорошо» (средний уровень); 85 – 100 % от максимально возможной суммы баллов – «отлично» (высокий (максимальный) уровень)				

3 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие процесс формирования компетенций в ходе освоения образовательной программы

3.1 Задания для текущего контроля успеваемости

Задания лабораторных работ

Тема: Основы анализа данных

Лабораторная работа «Начало работы в Deductor studio»

1. Создайте новый проект и сохраните его под именем dedst.ded. Не используйте упакованный формат файла.
2. Заполните свойства проекта.
3. Просмотрите файл проекта через любой текстовый редактор.
4. Сделайте видимой вкладку Подключения.
5. Поменяйте местами порядок вкладок Сценарии и Подключения.

6. Найдите в помощи раздел «Системные требования».
7. С помощью мастеров импорта, экспорта или обработки сформируйте сценарий.

Тема: Основы анализа данных
Лабораторная работа «Сценарии»

8. Создайте новый проект и сохраните его под именем test2.ded.
9. Создайте и сохраните в любом текстовом редакторе файл следующего вида:
10. a,1,4.5,b,c, 26/04/2007,d a1,0,5,b1,c1,,d1
11. Импортируйте его в Deductor, корректно настроив параметры импорта. Используйте относительный путь для файла. Метку узла переименуйте в Пример импорта файла. В комментарии к узлу впишите: Текстовый файл с разделителями-запятыми.
12. Добавьте к узлу узел Настройка набора данных и задайте следующие метки к столбцам: Поле1, Поле2, Поле3 и т.д.
13. Экпортируйте набор данных в текстовый файл с настройками, предлагаемыми по умолчанию.
14. Импортируйте только что экспортированный файл в Deductor.
15. Присоедините к новому узлу импорта (путем копирования) предыдущую ветвь, начиная с узла Настройка набора данных.
16. Между экспортом и настройкой набора данных вставьте еще один узел настройки, в котором измените тип столбца Поле2 на логический.
17. Удалите только что вставленный узел.
18. Сохраните проект.

Тема: Основы анализа данных
Лабораторная работа «Базовые визуализаторы»

- 1 В визуализаторе Таблица настройте, чтобы при отображении к значениям в Поле3 добавлялось слово «кг.». Сохраните конфигурацию визуализатора под названием «К1».
- 2 Сделайте первые три столбца невидимыми. Сохраните конфигурацию визуализатора под названием «К2».
- 3 В визуализаторе Таблица установите фильтр «Поле6 = не пустой». Удалите фильтр.
- 4 Создайте новый проект. Импортируйте в него текстовый файл CreditSample.txt.
- 5 Отсортируйте этот набор данных по следующим полям в порядке возрастания: Срок ссуды, Размер ссуды, Количество иждивенцев.
- 6 Сделайте следующую замену (после Сортировки) в поле Семейное положение: значение Да измените на Женат/замужем, Нет - на Холост/Не замужем.
- 7 Сделайте следующую замену (после предыдущего узла Замена данных) в поле Количество иждивенцев: значение 0 - на Нет, 1 - без изменений, 2 и 3 - 2 и более. Используйте два способа - непосредственным вводом в мастере обработки и через файл таблицы соответствий. Файл подстановок предварительно создайте в любом текстовом редакторе, например, в Блокноте.
- 8 Старое поле Количество иждивенцев удалите из набора данных, а новое поле Количество иждивенцев_REPLACE переименуйте в Иждивенцы.
- 9 Отфильтруйте набор данных, по полю Иждивенцы так, чтобы в выходной набор попали только строки, у которых значение в поле Иждивенцы не равно Нет.
- 10 Отфильтруйте набор данных, по полю Иждивенцы так, чтобы в выходной набор попали только строки, у которых значение в поле Иждивенцы не равно
- 11 Продолжите фильтровать набор данных, полученный в п. 12. Наложите следующий фильтр, в который попадают все записи, удовлетворяющие условиям а либо условиям б: Размер ссуды - от 2000 до 5000, Цель ссуды - Покупка товара. Цель ссуды - Иное.

12 Отсортируйте последний набор данных по полю Код.

Тема: Основы анализа данных

Лабораторная работа «Калькулятор и скрипты»

1 Создайте новый проект. Импортируйте в него текстовый файл *CreditSample.txt*, 2 Создайте новое поле *Дата обработки*, значения в котором равны текущей дате.

3 Создайте новое поле *Размер ссуды, у.е.*, который рассчитывается делением на 70 поля *Размер ссуды, руб.* Все значения в новом поле должны быть округлены до второго знака.

4 Создайте новое поле *Флаг*, значение в котором истинно, если выполняется условие: *Среднемесячный доход* > 2000 и *Наличие недвижимости* = Да.

5 Создайте еще один столбец, значение в котором равно 1, если выполняется условие: *Флаг* = TRUE и *Давать кредит* = FALSE.

6 Создайте новое поле *RATE*, в котором хранится значение в поле *Срок ссуды*, возведенное в степень 0,6.

7 Создайте новое поле *Сегмент*, которое делит всех заемщиков на сегменты по следующим правилам (используйте функцию **IF/IFF**): 1) ЕСЛИ *Количество лет* >= 50 и *Среднемесячный доход* < 6000 ТО *Сегмент* = Сегмент 1 2) ЕСЛИ *Количество лет* < 30 и ТО *Сегмент* = Сегмент 2 3) *Сегмент* = Сегмент 3 во всех остальных случаях, не удовлетворяющим п.п. 1) и 2).

9 Создайте новый проект. Импортируйте в него текстовый файл *Trade.txt*.

10 Добавьте после узла импорта 2-3 обработчика из изученных ранее.

11 Импортируйте в него текстовый файл *TradeSales.txt*. Добавьте к нему поле *Номер строки* (используйте функцию калькулятора RowNum()).

12 Добавьте к набору данных **скрипт**, выполняющий те же действия с набором данных, выполните его.

Тема: Основы анализа данных

Лабораторная работа «Визуализация»

1) Импортируйте в проект файл *fuel.txt*.

2) Создайте многомерную диаграмму:

- в виде поверхности;

- в виде топографической диаграммы.

3) Создайте диаграмму размещения.

4) Импортируйте в проект файл *CreditSamples.txt*.

а) Создайте гистограмму, иллюстрирующую распределение заемщиков по размеру ссуды.

б) Создайте гистограммы, иллюстрирующие распределение заемщиков по размеру ссуды с учетом их семейного положения. Сделайте вывод о влиянии семейного положения на размер ссуды.

5) Создайте кубы двух видов:

а) Отображение цен на топливо в осях Города – Топливо.

б) Отображение цен на топливо, используя выражения в калькуляторе вида:

• (Факт 1) Цена.Sum.Total.Топливо

• (Факт 2) Цена.Sum.Total.Города.

6) Импортируйте в проект файл *fuel_orkuga.txt*.

7) Создайте куб в осях Федеральные округа, Город – Топливо. Проанализируйте средние цены на топливо. Сделайте выводы об уровне цен на топливо в различных федеральных округах.

Тема: Корреляционно-регрессионный анализ

Лабораторная работа «Парная корреляция»

Известны данные по регионам России за 2019 год о среднемесячной номинальной начисленной заработной плате работников и прожиточном минимуме, рублей.

- 1) Постройте поле корреляции и сформулируйте гипотезу о форме связи.
- 2) Рассчитайте параметры уравнений парных регрессий:
 - линейной;
 - степенной;
 - логарифмической;
 - показательной.
- 3) Для каждой из регрессий:
 - вычислите средние ошибки аппроксимации, сделайте выводы о качестве эконометрических моделей;
 - совместите на одном графике поле корреляции и линию регрессии;
 - оцените тесноту связи с помощью показателей корреляции и детерминации;
 - оцените с помощью F-критерия Фишера статистическую надёжность результатов регрессионного моделирования;
 - выберите лучшее уравнение регрессии и дайте его обоснование.
- 4) Сделайте выводы.

Тема: Корреляционно-регрессионный анализ

Лабораторная работа «Прогноз значений результативного признака»

Известны данные по регионам России о среднемесячной номинальной начисленной заработной плате работников и прожиточном минимуме, рублей.

- 1) Постройте поле корреляции, рассчитайте параметры уравнения линейной регрессии.
- 2) Используя уравнение линейной регрессии, дайте точечный прогноз значений результативного признака y_i для каждого значения факторного признака x_i .
- 3) С доверительной вероятностью 90% (уровень значимости $\alpha=0,1$) дайте интервальный прогноз значений результативного признака $y_{объясн}$ для каждого значения факторного признака $x_o = x_i$.
- 4) Нанесите точечный и интервальный прогнозы на поле корреляции.

Тема: Множественная (многофакторная) регрессия и корреляция.

Лабораторная работа «Множественная регрессия»

Изучается влияние стоимости основных и оборотных средств на величину валового дохода торговых предприятий. Для этого по 10 торговым предприятиям были получены данные.

- 1) Определите парные коэффициенты корреляции.
- 2) Сделайте вывод о силе связи результативного и факторных признаков, о связи факторных признаков между собой.
- 3) Постройте линейное уравнение множественной регрессии и поясните экономический смысл его параметров.
- 4) Дайте оценку полученного уравнения на основе коэффициента детерминации и общего F-критерия Фишера.
- 5) Используя t-критерий Стьюдента, оцените значимость параметров уравнения регрессии.

Тема: Множественная (многофакторная) регрессия и корреляция.

Лабораторная работа «Фиктивные переменные»

В таблице 1 представлены данные о рынке строящегося жилья.

- 1) Постройте линейное уравнение регрессии, характеризующее зависимость цены от всех заданных факторов. Установите, какие факторы мультиколлинеарны.
- 2) Определите факторы, формировавшие цену квартир в строящихся домах. Сгенерируйте фиктивную переменную, отражающую местоположение квартиры и позволяющую разделить всю совокупность квартир на две группы: квартиры на севере города (районы под цифрами 1 и 2) и на юге города (районы под цифрами 3 и 4).
- 3) Отберите информативные факторы. Постройте уравнение регрессии со статистически значимыми факторами.
- 4) Оцените статистическую значимость полученного уравнения регрессии.

Тема: Бинарная классификация.

Лабораторная работа «Бинарная классификация»

При проверке медицинской диагностической системы, основанной на бинарном классификаторе, получены следующие результаты (по вариантам)

№	Состояние пациента	Предположение классификатора
1	здоров	болен
2	болен	болен
3	здоров	здоров
4	болен	здоров
5	болен	болен
6	здоров	здоров
7	здоров	болен
8	болен	здоров
9	здоров	здоров
10	болен	болен

1 Подготовка:

1.1 Подготовьте выборку данных в ПО *Microsoft Excel*.

1.2 Постройте диаграмму, отображающую выборку данных.

2 2 Классификация:

2.1 Задайте целевую функцию.

2.2 Определите коэффициенты функции гипотезы с помощью инструмента «Поиск решения».

2.3 Рассчитайте значения точности, чувствительности, F-критерия.

3 Сделайте вывод об эффективности этого классификатора.

Тема: Кластерный анализ.

Лабораторная работа «Кластеризация»

Имеются данные по нескольким объектам, характеризующихся двумя свойствами (по вариантам)

Объект	Св-во X	Св-во Y
1	19	1
2	2	5
3	46	9
4	4	3
5	21	11
6	46	5
7	31	3
8	27	1
9	16	9
10	55	12

11	33	5
12	15	6
13	8	6
14	9	4

1) Используя метод К-средних распределить объекты по двум кластерам с помощью программного обеспечения Excel или Mathcad.

2) Построить графики (диаграммы), иллюстрирующие результаты последовательных шагов расчётов и окончательный результат распределения объектов по кластерам.

Лабораторная работа «Карты Кохонена» (реализуется в форме практической подготовки).

1. Получить у преподавателя данные о клиентах некоторого банка.

2 Выполните построение карт Кохонена. Проанализируйте результаты, что можно сказать о вероятности возврата кредита для групп 2, 3 и 4?

3 Используя различные отображения карты Кохонена, постройте 3-4 правила выдачи кредитов.

Комплект заданий для расчётно-графической работы

«Анализ рынка недвижимости»

Задание:

Используя информацию сайтов о продаже квартир в г. Комсомольске-на-Амуре, г. Хабаровске сформировать массив данных по стоящим на продаже 120 квартирам (по вариантам):

- адрес квартиры,
- общая площадь квартиры в кв.м;
- стоимость квартиры в тыс. руб.

Используя сформированный массив данных:

- 1) Построить график корреляционного поля.
- 2) Сформулировать гипотезу о форме связи.
- 3) Рассчитать параметры четырёх различных уравнений регрессии.
- 4) Совместить график корреляционного поля с каждым из графиков уравнений регрессии.
- 5) Проверить значимость каждого из уравнений регрессии.
- 6) Из полученных уравнений регрессии выбрать наиболее подходящее.
- 7) Используя наиболее подходящее уравнение регрессии, дать точечный прогноз ожидаемой стоимости квартиры площадью 100 кв. м.
- 8) Оценить, в какой степени различия в стоимости квартир обусловлены различием их общей площади.