



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт математики и информационных технологий
Кафедра алгебраических и информационных систем



Рабочая программа дисциплины (модуля)

Б1.В.03 Интеллектуальный анализ данных

Направление подготовки	09.03.03 Прикладная информатика
Направленность (профиль) подготовки информационных систем	Проектирование и разработка информационных систем
Квалификация выпускника	бакалавр
Форма обучения	очная

Иркутск 2026 г.

I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ (МОДУЛЯ)

Цели: Целью курса является изучение современных подходов, моделей, алгоритмов анализа данных и решения задач распознавания, классификации, нахождения зависимостей.

Задачи:

- изучение сущности и роли анализа информации;
- ознакомление бакалавров с основными средствами и методами анализа данных при исследовании сложных систем;
- знакомство с основными технологиями анализа данных;
- формирование умений и навыков в области поддержки принятия решений с применением современных методов и средств анализа данных.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина (модуль) «Интеллектуальный анализ данных» относится к части, формируемой участниками образовательных отношений «Блок 1. Дисциплины (модули)».

Дисциплина изучается на третьем курсе.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы знания, умения и навыки, формируемые предшествующими дисциплинами:

- Математика;
- Дискретная математика;
- Теория вероятностей и математическая статистика;
- Языки программирования;
- Дискретные структуры.

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной:

- Системы искусственного интеллекта;
- Семантический искусственный интеллект;
- Методы машинного обучения.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование компетенций (элементов следующих компетенций) в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки:

**Перечень планируемых результатов обучения по дисциплине (модулю),
соотнесенных с индикаторами достижения компетенций**

Компетенция	Индикаторы компетенций	Результаты обучения
ПК11 Способен подготавливать и обрабатывать данные для аналитических исследований	ПК11.1 Очищает, нормализует и преобразует данные для анализа	• Знает современные проблемы анализа данных, теории распознавания, классификации, поиска зависимостей; методы и подходы решения практических задач анализа данных и классификации коллективами алгоритмов; программные средства решения основных задач анализа данных и классификации. • Умеет пользоваться своими знаниями для решения фундаментальных, прикладных и технологических задач в различных предметных областях; эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов. • Владеет навыками анализа большого объема частично противоречивых и неполных признаков описаний; навыками грамотной обработки статистических многомерных данных, оформления результатов численных расчетов и их сопоставления с теоретическими оценками;

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов, в том числе 36 часов на контроль, из них 36 часов на экзамен.

Из них реализуется с использованием электронного обучения и дистанционных образовательных технологий 4 часа самостоятельной работы.

Форма промежуточной аттестации: экзамен.

4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и СРС, отведенного на них количества академических часов

п/п	Раздел дисциплины/темы	Семестр	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости; Форма промежуточной аттестации (по семестрам)
			Контактная работа преподавателя с обучающимися			Самостоятельная работа	
			Лекции	Семинарские (практические) занятия	Консультации		
1	Подготовка данных	5	6	6	0	4	
2	Линейная регрессия	5	4	4	0	0	
3	Логистическая регрессия	5	2	2	0	0	
4	Регуляризация	5	2	2	0	0	
5	Метод градиентного спуска	5	4	4	0	0	
6	Метод k ближайших соседей	5	2	2	0	0	
7	Метод опорных векторов	5	4	4	0	0	
8	Деревья решений и случайный лес	5	4	4	0	0	
9	Кластеризация	5	6	6	0	0	
Итого за 5 семестр			34	34	0	4	Экз (36)
Итого часов			34	34	0	4	

4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Затраты времени, час. (из них с применением ДОТ)		

5	Подготовка данных	Для овладения знаниями: чтение учебной литературы, чтение дополнительной литературы Для закрепления и систематизации знаний: работа с конспектом лекций Для формирования умений: решение задач	1-3 недели курса	4 (4)	Тест, Пз	Задачи в ИОС Домик
Общая трудоемкость самостоятельной работы по дисциплине (час)				4		
Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий (час)				4		
Бюджет времени самостоятельной работы, предусмотренный учебным планом для данной дисциплины (час)				4		

4.3 Содержание учебного материала

Трудоемкость дисциплины (з.е.)	3
Наименование основных разделов (модулей)	Подготовка данных Линейная регрессия Логистическая регрессия Регуляризация Метод градиентного спуска Метод k ближайших соседей Метод опорных векторов Деревья решений и случайный лес Кластеризация
Формы текущего контроля	Тест, практическое задание, лабораторная работа, конспект лекций, устный опрос
Форма промежуточной аттестации	Экзамен

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины (модуля)	Наименование семинаров, практических и лабораторных работ	Трудоемкость, час. (из них электронные часы)	Оценочные средства	Формируемые компетенции
1	1	Объекты Series и DataFrame библиотеки Pandas. Очистка данных. Агрегация. Первичный анализ и визуализация данных	6 (0)	Тест, ЛР	ПК11.1

№ п/п	№ раздела и темы дисциплины (модуля)	Наименование семинаров, практических и лабораторных работ	Трудоемкость, час. (из них электронные часы)	Оценочные средства	Формируемые компетенции
2	2	Линейная парная регрессия. Линейная множественная регрессия. Линейная регрессия	4 (0)	Тест, ЛР, КЛ, УО	ПК11.1
3	3	Логистическая регрессия	2 (0)	Тест, ЛР, УО	ПК11.1
4	4	Регуляризация	2 (0)	Тест, ЛР, УО	ПК11.1
5	5	Алгоритм градиентного спуска. Нормализация данных.. Алгоритмы градиентного спуска. Градиентный спуск	4 (0)	Тест, ЛР, УО	ПК11.1
6	6	Метод k ближайших соседей	2 (0)	Тест, ЛР, УО	ПК11.1
7	7	Метод опорных векторов. Ядерные функции	4 (0)	Тест, ЛР, УО	ПК11.1
8	8	Деревья решений. Случайный лес. Деревья решений и случайный лес	4 (0)	Тест, ЛР, УО	ПК11.1
9	9	Метод k-средних. Агломеративная кластеризация. Эвристические методы кластеризации. Кластеризация	6 (0)	Тест, ЛР, УО	ПК11.1

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение самостоятельной работы студентов

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
1	Подготовка данных	Библиотека numpy	ПК11	ПК11.1

4.4. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов всех форм и видов обучения является одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего профессионального образования. Согласно требованиям нормативных документов самостоятельная работа студентов является обязательным компонентом образовательного процесса, так как она обеспечивает закрепление получаемых на лекционных занятиях знаний путем приобретения навыков осмысления и расширения их содержания, навыков решения актуальных проблем формирования общекультурных и профессиональных компетенций, научно-исследовательской деятельности, подготовки к семинарам, лабораторным работам, сдаче зачетов и экзаменов. Самостоятельная работа студентов представляет собой совокупность аудиторных и внеаудиторных занятий и работ. Самостоятельная работа в рамках образовательного процесса в вузе решает следующие задачи:

- закрепление и расширение знаний, умений, полученных студентами во время аудиторных и внеаудиторных занятий, превращение их в стереотипы умственной и физической деятельности;
- приобретение дополнительных знаний и навыков по дисциплинам учебного плана;
- формирование и развитие знаний и навыков, связанных с научно-исследовательской деятельностью;
- развитие ориентации и установки на качественное освоение образовательной программы;
- развитие навыков самоорганизации;
- формирование самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации;
- выработка навыков эффективной самостоятельной профессиональной теоретической, практической и учебно-исследовательской деятельности.

Подготовка к лекции. Качество освоения содержания конкретной дисциплины прямо зависит от того, насколько студент сам, без внешнего принуждения формирует у себя установку на получение на лекциях новых знаний, дополняющих уже имеющиеся по данной дисциплине. Время на подготовку студентов к двухчасовой лекции по нормативам составляет не менее 0,2 часа.

Подготовка к практическому занятию. Подготовка к практическому занятию включает следующие элементы самостоятельной деятельности: четкое представление цели и задач его проведения; выделение навыков умственной, аналитической, научной деятельности, которые станут результатом предстоящей работы. Выработка навыков осуществляется с помощью получения новой информации об изучаемых процессах и с помощью знания о том, в какой степени в данное время студент владеет методами исследовательской деятельности, которыми он станет пользоваться на практическом занятии. Подготовка к практическому занятию нередко требует подбора материала, данных и специальных источников, с которыми предстоит учебная работа. Студенты должны дома подготовить к занятию 3–4 примера формулировки темы исследования, представленного в монографиях, научных статьях, отчетах. Затем они самостоятельно осуществляют поиск соответствующих источников, определяют актуальность конкретного исследования процессов и явлений, выделяют основные способы доказательства авторами научных работ ценности того, чем они занимаются. В ходе самого практического занятия студенты сначала представляют

найденные ими варианты формулировки актуальности исследования, обсуждают их и обосновывают свое мнение о наилучшем варианте. Время на подготовку к практическому занятию по нормативам составляет не менее 0,2 часа.

Подготовка к контрольной работе. Контрольная работа назначается после изучения определенного раздела (разделов) дисциплины и представляет собой совокупность развернутых письменных ответов студентов на вопросы, которые они заранее получают от преподавателя. Самостоятельная подготовка к контрольной работе включает в себя: — изучение конспектов лекций, раскрывающих материал, знание которого проверяется контрольной работой; повторение учебного материала, полученного при подготовке к семинарским, практическим занятиям и во время их проведения; изучение дополнительной литературы, в которой конкретизируется содержание проверяемых знаний; составление в мысленной форме ответов на поставленные в контрольной работе вопросы; формирование психологической установки на успешное выполнение всех заданий. Время на подготовку к контрольной работе по нормативам составляет 2 часа.

Подготовка к экзамену. Самостоятельная подготовка к экзамену схожа с подготовкой к зачету, особенно если он дифференцированный. Но объем учебного материала, который нужно восстановить в памяти к экзамену, вновь осмыслить и понять, значительно больше, поэтому требуется больше времени и умственных усилий. Важно сформировать целостное представление о содержании ответа на каждый вопрос, что предполагает знание разных научных трактовок сущности того или иного явления, процесса, умение раскрывать факторы, определяющие их противоречивость, знание имен ученых, изучавших обсуждаемую проблему. Необходимо также привести информацию о материалах эмпирических исследований, что указывает на всестороннюю подготовку студента к экзамену. Время на подготовку к экзамену по нормативам составляет 36 часов для бакалавров.

Формы внеаудиторной самостоятельной работы

Составление глоссария Цель самостоятельной работы: повысить уровень информационный культуры; приобрести новые знания; отработать необходимые навыки в предметной области учебного курса. Глоссарий — словарь специализированных терминов и их определений. Статья глоссария — определение термина. Содержание задания: сбор и систематизация понятий или терминов, объединенных общей специфической тематикой, по одному либо нескольким источникам. Выполнение задания: 1) внимательно прочитать работу; 2) определить наиболее часто встречающиеся термины; 3) составить список терминов, объединенных общей тематикой; 4) расположить термины в алфавитном порядке; 5) составить статьи глоссария: — дать точную формулировку термина в именительном падеже; — объемно раскрыть смысл данного термина. Планируемые результаты самостоятельной работы: способность студентов решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

Разработка проекта (индивидуального, группового) Цель самостоятельной работы: развитие способности прогнозировать, проектировать, моделировать. Проект — «ограниченное во времени целенаправленное изменение отдельной системы с установленными требованиями к качеству результатов, возможными рамками расхода средств и ресурсов и специфической организацией». Выполнение задания: 1) диагностика ситуации (проблематизация, целеполагание, конкретизация цели, форматирование проекта); 2) проектирование (уточнение цели, функций, задач и плана работы; теоретическое моделирование методов и

средств решения задач; детальная проработка этапов решения конкретных задач; пошаговое выполнение запланированных проектных действий; систематизация и обобщение полученных результатов, конструирование предполагаемого результата, пошаговое выполнение проектных действий); 3) рефлексия (выяснение соответствия полученного результата замыслу; определение качества полученного продукта; перспективы его развития и использования). Предполагаемые результаты самостоятельной работы: готовность студентов использовать знание современных проблем науки и образования при решении образовательных и профессиональных задач; готовность использовать индивидуальные креативные способности для оригинального решения исследовательских задач; — способность прогнозировать, проектировать, моделировать.

Информационный поиск Цель самостоятельной работы: развитие способности к проектированию и преобразованию учебных действий на основе различных видов информационного поиска. Информационный поиск — поиск неструктурированной документальной информации. Список современных задач информационного поиска: решение вопросов моделирования; классификация документов; фильтрация, классификация документов; проектирование архитектур поисковых систем и пользовательских интерфейсов; извлечение информации (аннотирование и реферирование документов); выбор информационно-поискового языка запроса в поисковых системах. Содержание задания по видам поиска: поиск библиографический — поиск необходимых сведений об источнике и установление его наличия в системе других источников. Ведется путем разыскания библиографической информации и библиографических пособий (информационных изданий); поиск самих информационных источников (документов и изданий), в которых есть или может содержаться нужная информация; — поиск фактических сведений, содержащихся в литературе, книге (например, об исторических фактах и событиях, о биографических данных из жизни и деятельности писателя, ученого и т. п.). Выполнение задания:

- 1) определение области знаний;
- 2) выбор типа и источников данных;
- 3) сбор материалов, необходимых для наполнения информационной модели;
- 4) отбор наиболее полезной информации;
- 5) выбор метода обработки информации (классификация, кластеризация, регрессионный анализ и т.д.);
- 6) выбор алгоритма поиска закономерностей;
- 7) поиск закономерностей, формальных правил и структурных связей в собранной информации;
- 8) творческая интерпретация полученных результатов.

Планируемые результаты самостоятельной работы: — способность студентов решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; готовность использовать знание современных проблем науки и образования при решении образовательных и профессиональных задач.

Разработка мультимедийной презентации Цели самостоятельной работы (варианты): — освоение (закрепление, обобщение, систематизация) учебного материала; — обеспечение контроля качества знаний; — формирование специальных компетенций, обеспечивающих возможность работы с информационными технологиями; — становление общекультурных компетенций. Мультимедийная презентация — представление содержания

учебного материала, учебной задачи с использованием мультимедийных технологий.

Выполнение задания:

1. Этап проектирования: — определение целей использования презентации; — сбор необходимого материала (тексты, рисунки, схемы и др.); — формирование структуры и логики подачи материала; — создание папки, в которую помещен собранный материал.

2. Этап конструирования: — выбор программы MS PowerPoint в меню компьютера; — определение дизайна слайдов; — наполнение слайдов собранной текстовой и наглядной информацией; — включение эффектов анимации и музыкального сопровождения (при необходимости); — установка режима показа слайдов (титульный слайд, включающий наименование кафедры, где выполнена работа, название презентации, город и год; содержательный — список слайдов презентации, сгруппированных по темам сообщения; заключительный слайд содержит выводы, пожелания, список литературы и пр.).

3. Этап моделирования — проверка и коррекция подготовленного материала, определение продолжительности его демонстрации.

Планируемые результаты самостоятельной работы: — повышение информационной культуры студентов и обеспечение их готовности к интеграции в современное информационное пространство; — способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; — способность к критическому восприятию, обобщению, анализу профессиональной информации, постановке цели и выбору путей ее достижения; — способность применять современные методики и технологии организации и реализации образовательного процесса на различных образовательных ступенях в различных образовательных учреждениях; — готовность использовать индивидуальные креативные способности для оригинального решения исследовательских задач.

В ФБГОУ ВО «ИГУ» организация самостоятельной работы студентов регламентируется Положением о самостоятельной работе студентов, принятым Ученым советом ИГУ 22 июня 2012 г.

4.5. Примерная тематика курсовых работ (проектов)

По данной дисциплине выполнение курсовых проектов (работ) не предусматривается.

V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

а) основная литература

1. Сапрыкин, О. Н. Интеллектуальный анализ данных [Электронный ресурс] / О. Н. Сапрыкин. - Электрон. текстовые дан. - Самара : СамГУ, 2020. - 80 с. - ЭБС "Лань". - Неогранич. доступ. - ISBN 978-5-7883-1563-8 : Б. ц.

2. Грас, Джоэл. Data Science. Наука о данных с нуля [Текст] : учеб. пособие / Д. Грас. - СПб. : БХВ - Петербург, 2019. - 336 с. ; 23 см. - ISBN 978-5-9775-3758-2 : 790.02 р.

б) дополнительная литература

1. Цильковский, И. А. Методы анализа знаний и данных [Электронный ресурс] : уч.-методич. пособие / И. А. Цильковский. - Электрон. текстовые дан. - Новосибирск : Изд-во НГТУ, 2010. - 68 с. ; нет. - ЭБС "Руконт". - неогранич. доступ. - ISBN 978-5-7782-1377-7 : Б. ц.

2. Богданов, Е. П. Интеллектуальный анализ данных [Электронный ресурс] :

практикум для подготовки магистрантов направления 09.04.03 «прикладная информатика» профиль подготовки «информационные системы и технологии корпоративного управления» / Е. П. Богданов. - Электрон. текстовые дан. - Волгоград : Волгоградский ГАУ, 2019. - 112 с. - ЭБС "Лань". - Неогранич. доступ. - Б. ц.

в) периодическая литература

Нет.

г) базы данных, информационно-справочные и поисковые системы

Нет.

В соответствии с п. 4.3.4. ФГОС ВО, обучающимся в течение всего периода обучения обеспечен неограниченный доступ (удаленный доступ) к электронно-библиотечным системам:

— Открытая электронная база ресурсов и исследований «Университетская информационная система РОССИЯ» [Электронный ресурс] : сайт. – Режим доступа: <http://uisrussia.msu.ru> бессрочный

— Государственная информационная система «Национальная электронная библиотека» [Электронный ресурс] : сайт. – Режим доступа: <http://нэб.рф>. бессрочный

— Научная электронная библиотека «ELIBRARY.RU» [Электронный ресурс] : сайт. - Контракт № 148 от 23.12.2020 г. Акт от 24.12.2020 г. Срок действия по 31.12.2022 г. – Режим доступа: <http://elibrary.ru/>

— ЭБС «Издательство Лань». Контракт № 04-Е-0346 от 12.11.2021 г. № 976 от 14.11.2021 г. Срок действия по 13.11.2022 г. – Режим доступа: <https://www.e.lanbook.com>

— ЭБС ЭЧЗ «Библиотех». Государственный контракт № 019 от 22.02.2011 г. ООО «Библиотех». Лицензионное соглашение к Государственному контракту № 019 от 22.02.2011. Срок действия: бессрочный. – Режим доступа: <https://isu.bibliotech.ru/>

— ЭБС «Рукопт» ЦКБ «Бибком». № 04-Е-0343 от 12.11.2021 г. Акт № 6К-5195 от 14.11.2021 г. Срок действия по 13.11.2022г. – Режим доступа: <http://rucont.ru>

— ЭБС «Айбукс.ру/ibooks.ru» ООО «Айбукс». Контракт № 04-Е-0344 от 12.11.2021 г.; Акт от 14.11.2021 г. Срок действия по 13.11.2022 г. – Режим доступа: <http://ibooks.ru>

— Электронно-библиотечная система «ЭБС Юрайт». ООО «Электронное издательство Юрайт». Контракт № 04-Е-0258 от 20.09.2021г. Контракт № 04-Е-0258 от 20.09.2021 г. Срок действия по 17.10. 2022 г. – Режим доступа: <https://urait.ru>

— УБД ИВИС. Контракт № 04-Е-0347 от 12.11.2021 г. Акт от 15.11.2021 г. Срок действия с 01.01.2022 по 31.12.2022 г. – Режим доступа: <http://dlib.eastview.com>

— Электронная библиотека ИД Гребенников. Контракт № 04-Е-0348 от 12.11.2021г.; Акт № 348 от 15.11.2021 г. Срок действия с 01.01.2022 по 31.12.2022 – Режим доступа: <http://grebennikon.ru>

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
---	---	--

Специальные помещения: Учебная аудитория для проведения занятий лекционного и семинарского типа, текущего контроля, промежуточной аттестации.	Аудитория оборудована специализированной учебной мебелью, техническими средствами обучения, служащими для представления информации большой аудитории: Ноутбук(AserAspirev3-5516 (AMDA10-4600M 2300 МГц)) (1 штука) с неограниченным доступом к сети Интернет; Проектор Vivitek, экран ScreenVtdiaEcot- 3200*200MW 1:1, колонки, наборы демонстрационного оборудования и учебно-наглядных пособий, обеспечивающие тематические иллюстрации, соответствующие рабочей программе дисциплины «Архитектурный подход к развитию предприятий и информационных систем». Учебная лаборатория: компьютеры для проведения практических работ (Системный блок AMDAthlon-64 X3 445 3100 МГц), Монитор LG F1742S (2 штуки), Монитор ViewSonic VA703b(24 штуки) с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации; проектор Sony XGA VPLSX535, экран ScreenVtdiaEcot- 3200*200MW 1:1	ОС Windows: DreamSpark Premium, Договор № 03-016-14 от 30.10.2014 Microsoft Office: 0365ProPiusOpenStudents ShrdSvr ALNG subs VL NL I MthAcadmsStdnt w/Faculty (15000 лицензий) Kaspersky Endpoint Security длябизнеса- стандартный Russian Edition. 15002499 Node 1 year Educational License № 1B08-170221-054045-730-177 BusinessStudio Лицензия № 7464 (бессрочно)
--	---	---

Специальные помещения: компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, курсового проектирования (выполнения курсовых работ), организации самостоятельной работы, в том числе, научно-исследовательской	Аудитория оборудована специализированной учебной мебелью, техническими средствами обучения: компьютеры (системный блок AMD Athlon 64 X2 DualCore 3600+ 1900 МГц (15 штук), Монитор LGFlatron L1742SE (14 штук), Монитор ViewSonic VG720) с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.	ОС Windows: DreamSpark Premium, Договор № 03-016-14 от 30.10.2014 Microsoft Office: 0365ProPiusOpenStudents ShrdSvr ALNG subs VL NL I MthAcadmsStdnt w/Faculty (15000 лицензий) Kaspersky Endpoint Security для бизнеса- стандартный Russian Edition. 15002499 Node 1 year Educational License № 1B08-170221-054045-730-177
--	---	---

6.2. Программное обеспечение

№	Наименование Программног о продукта	Кол-во	Обоснование для пользования ПО	Дата выдачи лицензии	Срок действия права пользования
1	Интерпретатор CPython	Условия правообладателя	Условия правообладателя	Условия правообладателя	Условия правообладателя

6.3. Технические и электронные средства

Методической системой преподавания предусмотрено использование технических и электронных средств обучения и контроля знаний студентов: мультимедийные презентации, фрагменты фильмов.

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При реализации программы данной дисциплины используются различные образовательные технологии, в том числе электронное обучение и дистанционные образовательные технологии.

Проблемное обучение	Создание в учебной деятельности проблемных ситуаций и организация активной самостоятельной деятельности учащихся по их разрешению, в результате чего происходит творческое овладение знаниями, умениями, навыками, развиваются мыслительные способности
Разноуровневое обучение	У преподавателя появляется возможность помогать слабому, уделять внимание сильному, реализуется желание сильных учащихся быстрее и глубже продвигаться в образовании. Сильные учащиеся утверждают в своих способностях, слабые получают возможность испытывать учебный успех, повышается уровень мотивации учения.

Проектные методы обучения	Работа по данной методике дает возможность развивать индивидуальные творческие способности учащихся, более осознанно подходить к профессиональному и социальному самоопределению
Исследовательские методы в обучении	Дает возможность учащимся самостоятельно пополнять свои знания, глубоко вникать в изучаемую проблему и предполагать пути ее решения, что важно при формировании мировоззрения. Это важно для определения индивидуальной траектории развития каждого обучающегося
Лекционно-семинарско-зачетная система	Данная система дает возможность сконцентрировать материал в блоки и преподносить его как единое целое, а контроль проводить по предварительной подготовке обучающихся
Информационно-коммуникационные технологии	Изменение и неограниченное обогащение содержания образования, использование интегрированных курсов, доступ в ИНТЕРНЕТ.

Наименование тем занятий с использованием активных форм обучения:

№	Тема занятия	Вид занятия	Форма / Методы интерактивного обучения	Кол-во часов (из них электронные часы)
1				
2				
3				
4				
5				
6				

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8.1. Оценочные средства текущего контроля

№ п\п	Вид контроля	Контролируемые темы (разделы)	Компетенции, компоненты которых контролируются
-------	--------------	-------------------------------	--

1	Тест	Подготовка данных. Линейная регрессия. Логистическая регрессия. Регуляризация. Метод градиентного спуска. Метод k ближайших соседей. Метод опорных векторов. Деревья решений и случайный лес. Кластеризация.	ПК11.1
2	Практическое задание	Подготовка данных.	ПК11.1
3	Лабораторная работа	Подготовка данных. Линейная регрессия. Логистическая регрессия. Регуляризация. Метод градиентного спуска. Метод k ближайших соседей. Метод опорных векторов. Деревья решений и случайный лес. Кластеризация.	ПК11.1
4	Конспект лекций	Линейная регрессия.	ПК11.1
5	Устный опрос	Линейная регрессия. Логистическая регрессия. Регуляризация. Метод градиентного спуска. Метод k ближайших соседей. Метод опорных векторов. Деревья решений и случайный лес. Кластеризация.	ПК11.1

Примеры оценочных средств для текущего контроля

Демонстрационный вариант теста

1. Задание открытой формы. Введите ответ.

Напишите название метода, который применяет агрегационные функции к набору данных

2. Задание с множественным выбором. Выберите 3 правильных ответа.

Какие из функций являются агрегационными

- a. Нахождение остатка от деления на 10 для каждого элемента столбца.
- b. Количество элементов, равных определенному значению.
- c. Минимум из неотрицательных значений.
- d. Нормализация элементов столбца.
- e. Нахождение дисперсии элементов столбца.

3. Задание с множественным выбором. Выберите 2 правильных ответа.

Имеется набор информации о продажах некоторого товара, включающая в себя ID точки продажи, дату, цену товара, объем продаж. Какие из действий потребуется выполнить, чтобы построить зависимость объема продаж от цены?

- a. Найти диапазон цен и разбить его на интервалы.
- b. Сгруппировать данные по интервалам цен.
- c. Найти средневзвешенную цену товара для каждой даты.
- d. Сгруппировать данные по датам.
- e. Найти средний ежедневный объем продаж для каждой точки продажи.
- f. Сгруппировать данные по цене.

4. *Задание на последовательность. Расположите в правильном порядке.*

В стандартной программе анализа данных встречаются следующие методы. Укажите последовательность их применения.

- a. fit
- b. predict
- c. train_test_split
- d. roc_auc_score

5. *Задание открытой формы. Введите ответ.*

Напишите название метода, который позволяет осуществить подбор параметров модели (обучение)

6. *Задание открытой формы. Введите числовой ответ.*

Сколько независимых переменных (факторов) имеется в модели парной регрессии

7. *Задание открытой формы. Введите числовой ответ.*

В модели линейной регрессии пять независимых переменных (факторов). Сколько параметров имеет эта модель?

8. *Задание с единичным выбором. Выберите один правильный ответ.*

Метод наименьших квадратов в задаче регрессии заключается в

- a. Разбиении множества наблюдений на квадраты с минимальной суммарной площадью.
- b. нахождении квадратичной функции, описывающей зависимость переменной y от факторов.
- c. подборе коэффициентов, минимизирующих сумму квадратов отклонений.
- d. Добавлении к целевой функции суммы квадратов параметров.

9. *Задание открытой формы. Введите ответ.*

Напишите название метода, который позволяет найти вероятности принадлежности объектов к целевому классу при решении задачи классификации.

10. *Задание с единичным выбором. Выберите один правильный ответ.*

Для чего используется сигмоидная функция в модели логистической регрессии?

- a. Для вычисления вероятности принадлежности объекта целевому классу.
- b. Сигмоида используется в качестве линии, разделяющей классы..
- c. Для вычисления расстояния от разделяющей гиперплоскости до объекта.
- d. Сигмоида используется для оценки качества предсказания.

11. *Задание с единичным выбором. Выберите один правильный ответ.*

Что происходит при увеличении количества параметров модели?

- a. До определенного момента качество предсказания повышается, потом обычно снижается.
- b. Всегда увеличивается качество предсказания.
- c. Качество предсказания повышается до определенного предела и дальше почти не изменяется.
- d. Качество предсказания снижается из-за переобучения.

12. Задание с множественным выбором. Выберите 1 правильный ответ.

Через точки с какими координатами обязательно проходит ROC-кривая?

- a. (0;0.5)
- b. (-1;0)
- c. (0;0)
- d. (-1;-1)
- e. (1;1)

13. Задание с единичным выбором. Выберите один правильный ответ.

Регуляризацией L1 называется

- a. суммы коэффициентов модели
- b. суммы модулей коэффициентов модели.
- c. суммы квадратов коэффициентов модели.
- d. одного максимального коэффициента модели

14. Задание с единичным выбором. Выберите один правильный ответ.

Регуляризацией L2 называется прием изменения целевой функции путем добавления к ней

- a. суммы коэффициентов модели
- b. суммы квадратов коэффициентов модели
- c. одного максимального коэффициента модели
- d. суммы модулей коэффициентов модели

15. Задание открытой формы. Введите ответ.

При вычислении коэффициента детерминации на обучающей выборке было получено значение 0.87, а на тестовой - 0.69. Какие выводы и рекомендации по дальнейшей работе можно сделать?

16. Задание с единичным выбором. Выберите один правильный ответ.

В чем заключается метод наискорейшего градиентного спуска?

- a. В выборе дальнейшего направления спуска с учетом антиградиента и предшествующего направления.
- b. Нахождении точки с минимумом целевой функции в направлении антиградиента
- c. Подборе шага спуска экспериментальным путем для увеличения скорости сходимости
- d. Минимизации числа шагов градиентного спуска.

17. Задание с единичным выбором. Выберите один правильный ответ.

Метод k ближайших соседей используется при решении задачи

- a. регрессии

- b. нормализации
- c. кластеризации
- d. классификации

18. Задание с единственным выбором. Выберите один правильный ответ.

При построении какого вида моделей анализа данных используется функция Джини?

- a. Модель логистической регрессии.
- b. Модель ближайших соседей.
- c. Модель опорных векторов.
- d. Двоичные решающие деревья.

19. Задание с единственным выбором. Выберите один правильный ответ.

Что происходит при увеличении количества деревьев в модели случайного леса?

- a. Качество предсказания повышается до определенного предела и дальше почти не изменяется.
- b. Всегда увеличивается качество предсказания.
- c. Качество предсказания снижается из-за переобучения.
- d. До определенного момента качество предсказания повышается, потом обычно снижается.

20. Задание на соответствие. Соотнесите элементы двух списков.

Установите соответствие между классом модели и ее описанием

- | | |
|---------------------------|---|
| 1. LogisticRegression | 1. Выполняет кластеризацию объектов на основе выбора центроидов. |
| 2. RandomForestClassifier | 2. Выполняет классификацию объектов на основе применения множества деревьев классификации |
| 3. KMeans | 3. Позволяет предсказать числовое значение неизвестного параметра |
| 4. LinearRegression | 4. Выполняет классификацию объектов на основе расстояния от разделяющей гиперплоскости |

21. Задание с единственным выбором. Выберите один правильный ответ.

Базисными векторами в методе опорных векторов в случае линейной неразделимости называются объекты расстояние от которых до разделяющей гиперплоскости с учетом знака

- a. строго равно d
- b. является наименьшим
- c. больше или равно d
- d. меньше или равно d

22. Задание с единственным выбором. Выберите один правильный ответ.

Функция называется ядерной, если

- a. она представима в виде скалярного произведения $\langle f(x), f(y) \rangle$
- b. она представима в виде суперпозиции сверточных функций
- c. она сама и ее первая производная непрерывны.
- d. она непрерывна и ограничена

23. Задание открытой формы. Введите ответ.

Какие приемы можно применить для создания нелинейных моделей регрессии и класс-

ификации, используя модели линейной или логистической регрессии и модели опорных векторов?

24. *Задание открытой формы. Введите ответ.*

Кратко запишите, в чем заключается ядерный трюк для модели опорных векторов.

25. *Задание с единичным выбором. Выберите один правильный ответ.*

В алгоритме кластеризации по методу k средних параметр k задает

- a. минимальный размер кластера
- b. количество кластеров
- c. минимальное расстояние между центрами кластеров
- d. средний размер кластера

26. *Задание с единичным выбором. Выберите один правильный ответ.*

В алгоритме агломеративной кластеризации на каждом шаге

- a. объект перебрасывается из одного кластера в другой.
- b. выполняется разделение одного кластера на два.
- c. выполняется объединение двух кластеров.
- d. один объект добавляется к одному из кластеров

8.2. Оценочные средства для промежуточной аттестации

Вопросы к экзамену:

1. Задачи анализа данных. Методы исследования.
2. Data mining. Сбор информации, виды признаков. Первичная обработка данных.
3. Табличное представление данных. Соединение таблиц. Агрегация.
4. Метрики средних значений. Построение усредненных зависимостей на примере кривой эластичности спроса.
5. Очистка и нормализация данных
6. Библиотека numpy. Работа с матрицами
7. Библиотека Pandas. Объекты Series и DataFrame
8. Библиотека matplotlib
9. Понятие модели анализа данных. Функция штрафа (потерь).
10. Обоснование выбора функции потерь. Метод максимального правдоподобия.
11. Задача линейной регрессии. Парная и множественная регрессия. Построение нелинейных признаков
12. Метод наименьших квадратов. Вывод формулы наименьших квадратов из метода максимального правдоподобия.
13. Матричная постановка задачи линейной регрессии. Вывод формулы линейной регрессии для метода наименьших квадратов.
14. Оценка качества линейной регрессии. Коэффициент детерминации.
15. Линейная регрессия. в библиотеке scikit-learn. Построение нелинейных признаков и визуализация.
16. Задача классификации. Линейная классификация. Классификация на основе расстояния до разделяющей гиперплоскости.
17. Сигмоидные функции их свойства. Вид производной. Применение сигмоидной функции в задаче линейной классификации.

18. Логарифмическая функция потерь и ее вывод из принципа максимального правдоподобия.
19. Логистическая регрессия. в библиотеке scikit-learn. Построение нелинейных признаков и визуализация.
20. Оценка качества предсказания. Матрица ошибок. Точность, чувствительность, специфичность, надежность.
21. ROC-кривая. Алгоритм построения ROC-кривой. Метрика ROC-AUC.
22. Проблемы сложных моделей анализа данных. Переобучение и мультиколлинеарность.
23. Регуляризация. Виды регуляризации. Графическое обоснование.
24. Подбор параметров для регуляризации. Регуляризация в библиотеке scikit-learn.
25. Численное решение задачи оптимизации. Общее описание метода градиентного спуска.
26. Выбор шага в методе градиентного спуска. Метод скорейшего спуска.
27. Обзор методов градиентного спуска. Метод сопряженных градиентов.
28. Применение метода градиентного спуска для решения задачи логистической регрессии.
29. Метод ближайших соседей. Общее описание.
30. Подбор параметра k в методе ближайших соседей. Снижение размерности пространства признаков с использованием сингулярного разложения матриц.
31. Метод опорных векторов. Постановка задачи оптимизации для случая линейной разделимости
32. Метод опорных векторов. Постановка задачи оптимизации для случая линейной неразделимости
33. Метод опорных векторов. Прямая и двойственная задача. Двойственные коэффициенты.
34. Реализация предсказания для метода опорных векторов.
35. Ядерные функции. Ядерный трюк.
36. Ядерные функции. РБФ.
37. Метод опорных векторов в библиотеке scikit-learn
38. Модель двоичного решающего дерева. Алгоритм построения и предсказания по дереву.
39. Функции для оценки неравномерности разбиения, используемые при построении двоичного решающего дерева.
40. Методы борьбы с переобучением, используемые в двоичных решающих деревьях.
41. Модель случайного леса. Методы рандомизации, используемые при построении деревьев. Выбор числа деревьев.
42. Модель случайного леса в библиотеке scikit-learn
43. Метод градиентного бустинга.
44. Кластеризация по методу k -средних.
45. Итеративный алгоритм поиска центроидов для метода k -средних.
46. Агломеративная кластеризация. Меры близости.
47. Подбор параметров моделей кластеризации. Правило локтя.

48. Эвристические алгоритмы кластеризации. Алгоритм DBSCAN.

Разработчики:

_____	_____	_____
(подпись)	доцент (занимаемая должность)	К.Д. Кириченко (инициалы, фамилия)

Программа составлена в соответствии с требованиями ФГОС ВО и учетом рекомендаций ПООП по направлению подготовки 09.03.03 «Прикладная информатика».

Программа рассмотрена на заседании кафедры алгебраических и информационных систем

Протокол № __ от «__» _____ 20__ г.

Зав. кафедрой

В.И. Пантелеев

Настоящая программа, не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.