



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования
**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ИГУ»)**

Институт математики и информационных технологий
Кафедра вычислительной математики и оптимизации



Рабочая программа дисциплины (модуля)

Б1.В.11 Интеллектуальные методы анализа данных

Направление подготовки	01.03.02 Прикладная математика и информатика
Направленность (профиль) подготовки	Системная и бизнес-аналитика
Квалификация выпускника	бакалавр
Форма обучения	очная

Иркутск 2024 г.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цели: формирование компетенций специалиста по направлению «Прикладная математика и информатика» в предметной области, связанной с решением задач по интеллектуальному анализу эколого-экономических данных в рамках изучения основ теории обучения машин, включая дискриминантный, кластерный и регрессионный анализ данных.

Задачи:

- ✓ формирование умений работы с данными в вычислительных системах на основе методов интеллектуального анализа данных и базовых навыков создания таких исследований;
- ✓ формирование владений основными понятиями современных систем анализа данных и поддержки принятия решений, математическими и алгоритмическими основами конструирования систем, принимающих решения на основе анализа неизвестных заранее данных;
- ✓ приобретение практических навыков использования современных технологий машинного обучения для решения математических задач в области интеллектуального анализа и моделирования эколого-экономических данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Учебная дисциплина Б1.В.11 Интеллектуальные методы анализа данных относится к части Блока 1 образовательной программы, формируемой участниками образовательных отношений.

Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами:

Б1.В.ДВ.02.01 Технологии программирования для анализа данных

Б1.О.25 Эконометрика

Б1.В.03 Прикладная экономическая статистика

Б1.В.04 Большие данные

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

Б1.В.14 Управление информатизацией предприятия

ФТД.02 Анализ и обработка данных при разработке цифровых продуктов

Б2.В.03(П) Научно-исследовательская работа

Б2.О.01(Пд) Преддипломная практика

Б3.01(Д) Выполнение и защита выпускной квалификационной работы

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и ОП ВО по направлению подготовки 01.03.02 Прикладная математика и информатика:

ПК-4 Способен управлять информацией в процессе разработки и реализации проектов разного уровня сложности.

В результате освоения дисциплины обучающийся должен знать:

- современное состояние исследований в области интеллектуальных методов обучения;
- принципы построения систем машинного обучения;
- модели представления и описания технологий интеллектуального анализа данных;

уметь:

- проводить анализ предметной области;
- определять назначение, выбирать методы и средства для построения систем интеллектуального анализа данных;
- строить системы интеллектуального анализа данных;

владеть:

- навыками использования аппарата сложного анализа данных
- методами классификации информации и идентификации информации
- навыками реализации алгоритмов интеллектуального анализа данных.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 4 зачетных ед., 144 час.

Форма промежуточной аттестации: экзамен, курсовая работа.

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

Раздел дисциплины / тема	Сем.	Виды учебной работы				Самост. работа	Формы текущего контроля; Формы промежут. аттестации
		Контактная работа преподавателя с обучающимися					
		Лекции	Лаб. занятия	Практ. занятия			
Организация исходного кода и работа с Jupyter Notebook (Jupyter Lab)	7	-	2	-	2	Опорный конспект	
Ключевые библиотеки и пакеты для анализа данных на языке Python (NumPy, Matplotlib, SciPy, Pandas)	7	-	8	-	2	Защита отчета по практической работе	
Метрические алгоритмы классификации	7	-	4	-	2	Защита отчета по практической работе	
Дискриминантный анализ	7	-	4	-	2	Защита отчета по практической работе	
Байесовская теория решений	7	-	4	-	2	Защита отчета по практической работе	
Регрессия и классификация. Обзор методов оптимизации.	7	-	8	-	4	Защита отчета по практической работе	
Метод опорных векторов	7	-	6	-	4	Защита отчета по практической работе	
Искусственные нейронные сети	7	-	8	-	4	Защита отчета по практической работе	

Решающие деревья и их ансамбли	7	-	6	-	4	Защита отчета по практической работе
Графические модели. Байесовские сети. Марковские сети	7	-	10	-	4	Защита отчета по практической работе
Итого (7 семестр):			60		30	экз., курс. раб.

4.2. План внеаудиторной самостоятельной работы обучающихся по дисциплине

Раздел дисциплины / тема	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самост. работы
	Вид самост. работы	Сроки выполнения	Затраты времени		
Организация исходного кода и работа с Jupyter Notebook (Jupyter Lab)	1. Освоение базовых возможностей языка программирования Python. 2. Реализация контрольных примеров программ	К окончанию выполнения практической работы по данной теме	2	Опорный конспект	ОЛ*-3 ДЛ**- 1,2 ИР***- 1,2,3
Ключевые библиотеки и пакеты для анализа данных на языке Python (NumPy, Matplotlib, SciPy, Pandas)	1. Освоение ключевых вычислительных пакетов SciPy и пакетов работы с данными NumPy, Pandas. 2. Реализация контрольных примеров программ и заданий.	К окончанию выполнения практической работы по данной теме	2	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Метрические алгоритмы классификации	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	2	Аналитический отчет (markdown)	ОЛ - 1, 2, 3 ДЛ - 4
Дискриминантный анализ	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	2	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3

Байесовская теория решений	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	2	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Регрессия и классификация. Обзор методов оптимизации.	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	4	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Метод опорных векторов	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	4	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Искусственные нейронные сети	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	4	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Решающие деревья и их ансамбли	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	4	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Графические модели. Байесовские сети. Марковские сети	1. Работа с пакетом машинного обучения Python (R). 2. Подготовка практической работы.	К окончанию выполнения практической работы по данной теме	4	Аналитический отчет (markdown)	ОЛ – 3 ДЛ - 1,2 ИР – 1,2,3
Общая трудоемкость самостоятельной работы (час.)			30		
Из них с использованием электронного обучения и дистанционных образовательных технологий (час.)			30		

*ОЛ – основная литература

** ДЛ – дополнительная литература

*** ИР – интернет - ресурсы

4.3. Содержание учебного материала

Тема 1. Организация исходного кода и работа с Jupyter Notebook (Jupyter Lab).

Содержание: системы контроля версий. Модули и дистрибуция исходного кода.

Организация проекта и периспользование кода. Обзор среды Jupyter Notebook. Обзор среды Jupyter Lab. Основные принципы работы с github. Формат разметки текста Markdown. Модули расширения Jupyter Notebook.

Тема 2. Ключевые библиотеки и пакеты для анализа данных на языке Python (NumPy, Matplotlib, SciPy, Pandas)

Базовые команды, библиотеки и модули Python . Работа с векторами и матрицами в библиотеке NumPy. Агрегирование данных с помощью библиотеки Pandas. Построение графиков с помощью пакета Matplotlib. Пакеты численных методов в библиотеке SciPy. Методы машинного обучения в библиотеке Scikit-Learn.

Тема 3. Основные понятия в области машинного обучения.

Объекты и признаки. Типы задач распознавания. Понятие алгоритма распознавания. Методы обучения. Функционал качества алгоритма. Эмпирический риск. Переобучение. Обобщающая способность. Скользящий контроль.

Тема 4. Метрические алгоритмы классификации.

Математические основы работы с матрицами и векторами. Понятие расстояния. Метод ближайшего соседа. Метод k ближайших соседей. Метод k взвешенных ближайших соседей. Метод парзеновского окна.

Тема 5. Дискриминантный анализ..

Линейный дискриминант Фишера. Методы снижения размерностей. Сингулярное разложение. Метод главных компонент.

Тема 6. Байесовская теория решений.

Оценка параметров вероятностной модели. Метод максимального правдоподобия. Максимизация апостериорной вероятности.

Тема 7. Регрессия и классификация. Обзор методов оптимизации.

Моделирование распознаваемого параметра. Моделирование наблюдаемой величины. Регрессия. Бинарная классификация. Линейная регрессия. Метод наименьших квадратов. Регуляризация по Тихонову. Байесовская линейная регрессия. Разреженная линейная регрессия. Модели классификации. Логистическая регрессия.

Тема 8. Метод опорных векторов.

Линейно разделимая выборка. Линейно неразделимая выборка. Ядра. Алгоритм Платта.

Тема 9. Искусственные нейронные сети.

Перцептрон. Метод обратного распространения ошибки. Гессиан для функции ошибки. Переобучение нейронных сетей. Сверточные нейронный сети.

Тема 10. Решающие деревья и их ансамбли.

Перцептрон. Метод обратного распространения ошибки. Гессиан для функции ошибки. Переобучение нейронных сетей. Сверточные нейронный сети.

Тема 11. Графические модели. Байесовские сети. Марковские сети..

Максимизация апостериорной вероятности для марковских цепей и деревьев. Минимизация энергии с помощью разрезов графов. Алгоритм а-расширения. Алгоритм распространения доверия. Структурный метод опорных векторов.

Тема 12. Практическое применение методов машинного обучения в области задач моделирования эколого-экономических систем

Моделирование показателей состояния и качества окружающей среды. Моделирование

показателей состояния и качества окружающей среды. Балансовые и оптимизационные эколого-экономические модели. Глобальные эколого-экономические модели. Динамические эколого-экономические модели. Моделирование истощаемых природных ресурсов.

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

Тема занятия	Всего часов	Оценочные средства	Формируемые компетенции
Организация исходного кода и работа с Jupyter Notebook (Jupyter Lab)	2	Опорный конспект	ПК-4
Ключевые библиотеки и пакеты для анализа данных на языке Python (NumPy, Matplotlib, SciPy, Pandas)	8	Аналитический отчет (markdown)	ПК-4
Метрические алгоритмы классификации	4	Аналитический отчет (markdown)	ПК-4
Дискриминантный анализ	4	Аналитический отчет (markdown)	ПК-4
Байесовская теория решений	4	Аналитический отчет (markdown)	ПК-4
Регрессия и классификация. Обзор методов оптимизации.	8	Аналитический отчет (markdown)	ПК-4
Метод опорных векторов	6	Аналитический отчет (markdown)	ПК-4
Искусственные нейронные сети	8	Аналитический отчет (markdown)	ПК-4
Решающие деревья и их ансамбли	6	Аналитический отчет (markdown)	ПК-4
Графические модели. Байесовские сети. Марковские сети	10	Аналитический отчет (markdown)	ПК-4

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы

Тема	Задание	Формируемые компетенции
Переобучение. Влияние размера обучающей выборки на качество обучения.	1. Изучение основных теоретических положений. 2. Подготовка опорного конспекта.	ПК-4
Переобучение. Влияние размера обучающей выборки на качество обучения.	1. Изучение основных модулей и функций библиотеки. 2. Подготовка опорного конспекта.	ПК-4
EM-алгоритм. Обучение с помощью EM-алгоритма для распределения Стюдента.	1. Изучение основных модулей и функций библиотеки. 2. Подготовка опорного конспекта. 3. Выполнение практического задания с помощью среды Jupyter Notebook (Python).	ПК-4
Метод опорных векторов. Искусственные нейронные сети. Распознавание символов текста с помощью метода опорных векторов и с помощью сверточных сетей.	1. Изучение основных модулей и функций библиотеки. 2. Подготовка опорного конспекта. 3. Выполнение практического задания с помощью среды Jupyter Notebook (Python).	ПК-4
Решающие деревья. Построение решающих деревьев. Алгоритм Gradient boosting.	1. Изучение основных модулей и функций библиотеки. 2. Подготовка опорного конспекта. 3. Выполнение практического задания с помощью среды Jupyter Notebook (Python).	ПК-4

Графические модели. Сегментация изображений с помощью графических моделей. Алгоритм Витерби для марковских цепей.	1. Изучение основных модулей и функций библиотеки. 2. Подготовка опорного конспекта. 3. Выполнение практического задания с помощью среды Jupyter Notebook (Python).	ПК-4
---	---	------

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 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 семинарскому занятию. Самостоятельная подготовка к семинару направлена: на развитие способности к чтению научной и иной литературы; на поиск дополнительной информации, позволяющей глубже разобраться в некоторых вопросах; на выделение при работе с разными источниками необходимой информации, которая требуется для полного ответа на вопросы плана семинарского занятия; на выработку умения правильно выписывать высказывания авторов из имеющихся источников информации, оформлять их по библиографическим нормам; на развитие умения осуществлять анализ выбранных источников информации; на подготовку собственного выступления по обсуждаемым вопросам; на формирование навыка оперативного реагирования на разные мнения, которые могут возникать при обсуждении тех или иных научных проблем. Время на подготовку к семинару по нормативам составляет не менее 0,2 часа.

Подготовка к коллоквиуму. Коллоквиум представляет собой коллективное обсуждение раздела дисциплины на основе самостоятельного изучения этого раздела студентами. Подготовка к данному виду учебных занятий осуществляется в следующем порядке. Преподаватель дает список вопросов, ответы на которые следует получить при изучении определенного перечня научных источников. Студентам во внеаудиторное время необходимо прочитать специальную литературу, выписать из нее ответы на вопросы, которые будут обсуждаться на коллоквиуме, мысленно сформулировать свое мнение по каждому из вопросов, которое они выскажут на занятии. Время на подготовку к коллоквиуму 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 зачету. Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия: перечитать все лекции, а также материалы, которые готовились к семинарским и практическим занятиям в течение семестра, соотнести эту информацию с вопросами, которые даны к зачету, если информации недостаточно, ответы находят в предложенной преподавателем литературе. Рекомендуется делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

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.

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

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) основная литература:

1. Алексеев, Д. С. Технологии интеллектуального анализа данных : учебник для вузов / Д. С. Алексеев, О. В. Щекочихин. — Санкт-Петербург : Лань, 2022. — 176 с. — ISBN 978-5-8114-8299-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/187559> (дата обращения: 05.05.2022). — Режим доступа: для авториз. пользователей.
2. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206711> (дата обращения: 05.05.2022). — Режим доступа: для авториз. пользователей.
3. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 210 с. — (Высшее образование). — ISBN 978-5-534-14638-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://www.ura.it.ru/bcode/492920> (дата обращения: 30.05.2022).

б) дополнительная литература:

1. Остроух, А. В. Интеллектуальные информационные системы и технологии : монография / А. В. Остроух, А. Б. Николаев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 308 с. — ISBN 978-5-8114-8578-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/177839> (дата обращения: 05.05.2022). — Режим доступа: для авториз. пользователей.
2. Сапрыкин, О. Н. Интеллектуальный анализ данных : учебное пособие / О. Н. Сапрыкин. — Самара : Самарский университет, 2020. — 80 с. — ISBN 978-5-7883-1563-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/188906> (дата обращения: 05.05.2022). — Режим доступа: для авториз. пользователей.
3. Воронов, М. В. Системы искусственного интеллекта: учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. — Москва : Издательство Юрайт, 2022. — 256 с. — (Высшее образование). — ISBN 978-5-534-14916-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://www.ura.it.ru/bcode/485440> (дата обращения: 30.05.2022).
4. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2022. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://www.ura.it.ru/bcode/490020> (дата обращения: 22.05.2022).

в) базы данных, информационно-справочные и поисковые системы:

1. Научная электронная библиотека «ELIBRARY.RU» [Электронный ресурс] : сайт. – Режим доступа: <http://elibrary.ru/defaultx.asp>
2. Открытая электронная база ресурсов и исследований «Университетская информационная система РОССИЯ» [Электронный ресурс] : сайт. – Режим доступа: <http://uisrussia.msu.ru>
3. Государственная информационная система «Национальная электронная библиотека» [Электронный ресурс] : сайт. – Режим доступа: <http://нэб.рф>
4. В соответствии с п. 4.3.4. ФГОС ВО, обучающимся в течение всего периода обучения обеспечен неограниченный доступ (удаленный доступ) к электронно-библиотечным системам:
 - ЭБС «Издательство Лань». ООО «Издательство Лань». Контракт № 92 от 12.11.2018 г. Акт от 14.11.2018 г.
 - ЭБС ЭЧЗ «Библиотех». Государственный контракт № 019 от 22.02.2011 г. ООО «Библиотех». Лицензионное соглашение № 31 от 22.02.2011 г. Адрес доступа: <https://isu.bibliotech.ru/> Срок действия: с 22.11.2011 г. бессрочный.
 - ЭБС «Национальный цифровой ресурс «Руконт». ЦКБ «Бибком». Контракт № 91 от 12.11.2018 г. Акт от 14.11.2018 г..
 - ЭБС «Айбукс.ру/ibooks.ru». ООО «Айбукс». Контракт № 90 от 12.11.2018 г. Акт № 54 от 14.11.2018 г.
 - Электронно-библиотечная система «ЭБС Юрайт». ООО «Электронное издательство Юрайт». Контракт № 70 от 04.10.2018 г.

е) интернет-ресурсы

1. <http://wombat.org.ua/AByteOfPython/AByteofPythonRussian-2.01.pdf> - это свободная книга по программированию на языке Python.
2. <https://www.python.org/> - язык программирования Python
3. www.gks.ru - Официальный сайт Государственного комитета статистики РФ
4. sophist.hse.ru - Единый архив экономических и социологических данных
5. http://library.hse.ru/e-resources/HSE_economic_journal - Экономический журнал Высшей школы экономики
6. <http://www.mnr.gov.ru/> - Министерство природных ресурсов и экологии РФ
7. <http://mpr.orb.ru/> - Министерство природных ресурсов, экологии и имущественных отношений Оренбургской области
8. www.cemi.rssi.ru - Центральный экономико-математический институт (ЦЭМИ) РАН
9. www.forecast.ru - Центр макроэкономического анализа и прогнозирования при ИНИ РАН
10. <http://www.biodat.ru/> - Информационный ресурс по биоразнообразию, экологическим рейтингам и др.
11. www.sci.aha.ru/ATL/ra00.htm - Web-атлас «Окружающая среда и здоровье населения России»
12. www.dmb.biophys.msu.ru/models - Динамические модели в биологии: реестр моделей.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование

ЭТОТ РАЗДЕЛ НЕ ЗАПОЛНЯТЬ

6.2. Программное обеспечение

1. Язык программирования Python 3.7. и выше (открытое программное обеспечение).
2. Язык программирования R 3.x и выше (открытое программное обеспечение).

3. R Studio Desktop 1.3.x и выше (открытое программное обеспечение).
4. Jupyter Notebook (открытое программное обеспечение).
5. Jupyter Lab (открытое программное обеспечение).
6. VisualStudioCode (Microsoft, открытое программное обеспечение)
7. Офисный пакет Microsoft Office Project Professional 2019 (лицензия ИГУ для образовательных учреждений).
8. Редакционно-издательская система MikTeX (открытое программное обеспечение).

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Оценочные средства текущего контроля

Вид контроля	Контролируемые темы	Контролируемые компетенции
Опорный конспект	Организация исходного кода и работа с Jupyter Notebook (Jupyter Lab)	ПК-4
Аналитический отчет	Ключевые библиотеки и пакеты для анализа данных на языке Python (NumPy, Matplotlib, SciPy, Pandas) Метрические алгоритмы классификации Дискриминантный анализ Байесовская теория решений Регрессия и классификация. Обзор методов оптимизации. Метод опорных векторов Искусственные нейронные сети Решающие деревья и их ансамбли Графические модели. Байесовские сети. Марковские сети Практическое применение методов машинного обучения в области задач моделирования эколого-экономических систем	ПК-4

Примеры оценочных средств текущего контроля

1. *Углубленные направления исследовательской работы в области методов машинного обучения*

Часть 1. Работа с глубокими сетями:

- Построение и обучение LSTM / GRU сетей в Keras
- Распознавание сущностей

Часть 2. Обработка изображений при помощи глубокого обучения:

- Сегментация изображений
- Детекция изображений

Часть 3. Проектирование интеллектуальных моделей для решения эколого-экономических задач:

- Моделирование показателей состояния и качества окружающей среды.
- Моделирование показателей состояния и качества окружающей среды.
- Динамические эколого-экономические модели.

- Моделирование истощаемых природных ресурсов.

2. Темы докладов

1. Определение логической модели представления знаний.
 2. Определение стратегии управления в экспертных системах и приведите классификацию стратегий.
 3. Формальное определение продукционной системы (по Е.Посту и А.Ньюэллу).
 4. Ведущие тенденции в разработке ИС для естественно-языковых систем.
 5. Главные направления структуризации при проектировании экспертных систем.
 6. Методы, применяемые для тестирования современных экспертных систем.
 7. Основные направления использования проблемно/предметно-ориентированных ИС.
 8. Основные различия между подходом, использующим управляемые образцами правила, и подходом, использующим управляемые образцами модули.
 9. Основные типы проблемных сред и ИС
 10. Примеры применения технологии динамических экспертных систем.
 11. Примеры успешного применения технологии ЭС.
 12. Назначение механизма инспекции в современных экспертных системах.
 13. Основные причины успеха современной технологии ЭС.
 14. Понятия интегрированного приложения, открытого приложения и распределенного приложения.
 15. Различия между рабочими пространствами и модулями прикладной системы.
 16. Статические и динамические задачи.
 17. Метод поиска решений в альтернативных пространствах при неполных и неточных данных.
 18. Метод поиска решений в иерархии пространств.
 19. Метод поиска решений в одном пространстве.
 20. Метод поиска решений с использованием нескольких моделей.
 21. Основные режимы работы ЭС.
 22. Базовые типы диаграмм, используемые в методологии разработки экспертных систем.
 23. Главные тенденции в разработке ИС для экспертных систем.
 24. Достоинства и особенности систем, управляемых образцами.
 25. Основные этапы разработки ЭС.
 26. Новые направления ИИ, в которых обозначились существенные практические успехи.
 27. Основные аспекты организации знаний в рабочей памяти и базе знаний ЭС.
 28. Основные компоненты статической ЭС.
 29. Основные параметры, определяющие свойства предметной области.
 30. Отличия архитектуры динамической ЭС от архитектуры статической ЭС.
- ## 3. Примеры практических заданий

Практическое задание № 1 (Линейная регрессия)

Ответьте на вопросы о данных по авиарейсам в США за январь-апрель 2008 года. По ссылке расположены Данные и их описание 1) Считайте выборку из файла при помощи функции `pd.read_csv` и ответьте на следующие вопросы: – Имеются ли в данных пропущенные значения? – Сколько всего пропущенных элементов в таблице "объект-

признак"? – Сколько объектов имеют хотя бы один пропуск? – Сколько признаков имеют хотя бы одно пропущенное значение? 2) Преобразуйте каждый признак FeatureName из указанных в пару новых признаков FeatureName_Hour, FeatureName_Minute, разделив каждое из значений на часы и минуты. Не забудьте при этом исключить исходный признак из выборки. В случае, если значение признака отсутствует, значения двух новых признаков, его заменяющих, также должны отсутствовать. 3) Некоторые из признаков, отличных от целевой переменной, могут оказывать чересчур значимое влияние на прогноз, поскольку по своему смыслу содержат большую долю информации о значении целевой переменной. Изучите описание датасета и исключите признаки, сильно коррелирующие с ответами. Ваш выбор признаков для исключения из выборки обоснуйте. 4) Приведите данные к виду, пригодному для обучения линейных моделей. Для этого вещественные признаки надо отмасштабировать, а категориальные — привести к числовому виду. Также надо устранить пропуски в данных. Реализуйте функцию transform_data, которая принимает на вход DataFrame с признаками и выполняет следующие шаги: – Замена пропущенных значений на нули для вещественных признаков и на строки 'nan' для категориальных. – Масштабирование вещественных признаков с помощью StandardScaler. – One-hot-кодирование категориальных признаков с помощью DictVectorizer или функции pd.get_dummies. Метод должен возвращать преобразованный DataFrame, который должна состоять из масштабированных вещественных признаков и закодированных категориальных (исходные признаки должны быть исключены из выборки). 5) Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30 (для этого можно использовать функцию train_test_split). 6) Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения MSE и R^2 и на этой подвыборке и контрольной выборке. Проинтерпретируйте полученный результат — насколько качественные прогнозы строит полученная модель? Какие проблемы наблюдаются в модели? 7) Обучите линейные регрессии с L1- и L2-регуляризатором, подобрав лучшее значение параметра регуляризации при помощи кросс-валидации. Выведите значения и на обучающей и контрольной выборках. Удалось ли решить указанные вами ранее проблемы?

Практическое задание 2 (работа с текстовыми данными и категориальными признаками и задачам бинарной классификации)

В рамках данного задания нужно решить задачу бинарной классификации для предсказания уровня заработной платы по тексту объявления о вакансии на примере набора данных с соревнования на Kaggle. Данные доступны по ссылке. 1) Разбейте получившуюся выборку на обучающую и контрольную в соотношении 70/30 2) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, осуществив векторизацию получившихся текстов при помощи CountVectorizer, обучив его на обучающей выборке и применив на тестовой. 3) Обучите логистическую регрессию из модуля sklearn с параметрами по умолчанию на обучающей выборке: 4) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на тестовой выборке. 5) Отсортируйте веса признаков для модели. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат. 6) Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, вычислив вектор tf-idf для каждого объекта помощи TfidfVectorizer, обучив его на обучающей выборке и применив на тестовой. 7) Заново обучите модель 8) Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на контрольной выборке. 9). Сравните значения метрик из п. 8 со значениями,

полученными в п. 4, и сравните соответствующие модели по качеству из работы. 11. Отсортируйте веса признаков для модели логистической регрессии из scikit-learn, полученной в п. 7. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.

7.2. Оценочные средства для промежуточной аттестации

Список вопросов для промежуточной аттестации:

Примерный перечень вопросов:

1. Работа с векторами и матрицами в библиотеке NumPy.
2. Агрегирование данных с помощью библиотеки Pandas
3. Построение графиков с помощью пакета Matplotlib
4. Пакеты численных методов в библиотеке SciPy.
5. Методы машинного обучения в библиотеке Scikit-Learn
6. Обзор задач, решаемых алгоритмами машинного обучения
7. Классификация алгоритмов машинного обучения
8. Линейные модели регрессии
9. Базисные функции
10. Регуляризация
11. Целевая функция логистической регрессии
12. Регуляризация логистической регрессии
13. Структура нейрона
14. Структура нейронной сети
15. Перцептрон
16. Обучения нейронной сети с помощью алгоритма обратного распространения ошибки
17. Структура деревьев решений
18. Виды разделяющих функций
19. Обучения дерева решений
20. Алгоритм Random Forest
21. Алгоритм AdaBoost
22. Каскад классификаторов
23. Кластеризация
24. Обучение без учителя
25. Алгоритм k-means
26. Иерархическая кластеризация
27. Моделирование показателей состояния и качества окружающей среды.
28. Моделирование показателей состояния и качества окружающей среды.
29. Балансовые и оптимизационные эколого-экономические модели,
30. Глобальные эколого-экономические модели.
31. Динамические эколого-экономические модели.
32. Моделирование истощаемых природных ресурсов.

Примеры заданий:

Тесты и задания в ЭИОС ИГУ на сайте <https://educa.isu.ru/>

Разработчик: Кедрин В.С., к.т.н., доцент, доцент