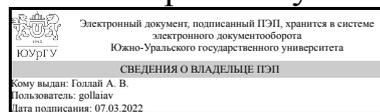


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Директор института
Высшая школа электроники и
компьютерных наук



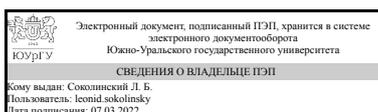
А. В. Голлай

РАБОЧАЯ ПРОГРАММА

**дисциплины 1.Ф.07 Интеллектуальный анализ данных
для направления 09.04.04 Программная инженерия
уровень Магистратура
форма обучения очная
кафедра-разработчик Системное программирование**

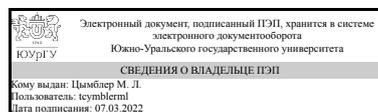
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



Л. Б. Соколинский

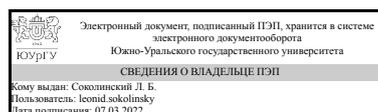
Разработчик программы,
д.физ.-мат.н., доц., профессор



М. Л. Цымблер

СОГЛАСОВАНО

Руководитель направления
д.физ.-мат.н., проф.



Л. Б. Соколинский

1. Цели и задачи дисциплины

Целью дисциплины является ознакомление магистрантов с основными методами и алгоритмами интеллектуального анализа данных.

Краткое содержание дисциплины

Введение в дисциплину. Феномен Больших данных. Понятие интеллектуального анализа данных. Технологический цикл анализа данных. Основные задачи интеллектуального анализа данных: поиск шаблонов, классификация, кластеризация, поиск аномалий. Поиск шаблонов. Понятия транзакции, частого набора, шаблона, поддержки, достоверности. Основные алгоритмы поиска частых наборов: Apriori, Eclat, FP-Growth. Выбор полезных шаблонов на основе мер support, confidence, lift и др. Компактное представление частых наборов: максимально частые и замкнутые наборы, иерархии наборов. Фрагментация и сэмплинг для поиска частых наборов. Классификация. Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Меры оценки доли примесей в узле дерева решений: индекс Джини, энтропия; алгоритмы классификации ID3, C4.5, CART. Байесовская классификация. Классификация по ближайшим соседям. Оценка качества классификации: меры Accuracy, Precision, Recall, F1. Ансамблевая классификация: бэггинг, бустинг, случайный лес. Кластеризация. Задачи кластеризации данных и подходы к ее решению. Разделительная кластеризация: алгоритмы k-means, k-medoids и др. Иерархическая кластеризация: дендрограммы, агломеративный и дивизимный подход. Меры схожести кластеров: Single linkage, Complete linkage, Group average и др. Плотностная кластеризация: алгоритм DBSCAN. Нечеткая кластеризация: алгоритм Fuzzy C-Means. Меры качества кластеризации: критерий Хопкинса, кросс-валидация, метод локтя, силуэтный коэффициент и др. Поиск аномалий. Понятия аномалии (выброса), шума, новизны в данных. Виды аномалий: точечные, глобальные, контекстные, смешанные. Статистические методы поиска аномалий: z-значимость, правило трех сигм, гистограммы. Поиск аномалий на основе расстояния. Поиск аномалий на основе плотности: метод вложенных циклов, метод решеток. Поиск аномалий с помощью разделительной и плотностной кластеризации. Поиск аномалий на основе классификации: метод One Class SVM, метод изолирующего леса.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-7 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач	Знает: определение базовых задач машинного обучения (поиск шаблонов, классификация, кластеризация и поиск аномалий) и основные алгоритмы их решения Умеет: определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.Ф.05 Разработка интеллектуальных систем на языке R, 1.О.10 Машинное обучение, 1.Ф.06 Глубокие нейронные сети	Не предусмотрены

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
1.Ф.05 Разработка интеллектуальных систем на языке R	Знает: методы и критерии оценки качества моделей машинного обучения Умеет: выбирать, применять и интегрировать методы и инструментальные средства систем искусственного интеллекта, критерии их выбора и методы комплексирования в рамках создания интегрированных гибридных интеллектуальных систем различного назначения Имеет практический опыт:
1.Ф.06 Глубокие нейронные сети	Знает: принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения (с подкреплением и без) Умеет: ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения Имеет практический опыт:
1.О.10 Машинное обучение	Знает: классы методов и алгоритмов машинного обучения, основные методы машинного обучения и их готовые реализации в библиотеке sklearn языка Python Умеет: реализовывать алгоритмы машинного обучения и производить их оптимальную настройку Имеет практический опыт: анализа, оптимизации и валидации моделей машинного обучения

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 74,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	144	144
<i>Аудиторные занятия:</i>	64	64
Лекции (Л)	32	32
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0

Самостоятельная работа (СРС)	69,5	69,5
с применением дистанционных образовательных технологий	0	
Индивидуальное задание	69,5	69,5
Консультации и промежуточная аттестация	10,5	10,5
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Введение в дисциплину	2	2	0	0
2	Поиск шаблонов	16	8	8	0
3	Классификация	16	8	8	0
4	Кластеризация	16	8	8	0
5	Поиск аномалий	14	6	8	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Феномен Больших данных. Понятие интеллектуального анализа данных. Технологический цикл анализа данных. Основные задачи анализа данных: поиск шаблонов, классификация, кластеризация, поиск аномалий.	2
2	2	Понятия транзакции, частого набора, шаблона, поддержки, достоверности. Основные алгоритмы поиска частых наборов: Apriori, Eclat, FP-Growth.	4
3	2	Выбор полезных шаблонов на основе мер support, confidence, lift и др. Компактное представление частых наборов: максимально частые и замкнутые наборы, иерархии наборов. Фрагментация и сэмплинг для поиска частых наборов	4
4	3	Процесс классификации: обучение модели, оценка модели, применение модели. Деревья решений. Меры оценки доли примесей в узле дерева решений: индекс Джини, энтропия; алгоритмы классификации ID3, C4.5, CART. Байесовская классификация. Классификация по ближайшим соседям. Оценка качества классификации: меры Accuracy, Precision, Recall, F1.	6
5	3	Ансамблевая классификация: бэггинг, бустинг, случайный лес.	2
6	4	Задачи кластеризации данных и подходы к ее решению. Разделительная кластеризация: алгоритмы k-means, k-medoids и др.	3
7	4	Иерархическая кластеризация: дендрограммы, агломеративный и дивизимный подход. Меры схожести кластеров: Single linkage, Complete linkage, Group average и др. Плотностная кластеризация: алгоритм DBSCAN. Нечеткая кластеризация: алгоритм Fuzzy C-Means.	3
8	4	Меры качества кластеризации: критерий Хопкинса, кросс-валидация, метод локтя, силуэтный коэффициент и др.	2
9	5	Понятия аномалии (выброса), шума, новизны в данных. Виды аномалий: точечные, глобальные, контекстные, смешанные. Статистические методы поиска аномалий: z-значимость, правило трех сигм, гистограммы. Поиск аномалий на основе расстояния. Поиск аномалий на основе плотности: метод вложенных циклов, метод решеток. Поиск аномалий с помощью разделительной и плотностной кластеризации. Поиск аномалий на основе классификации: метод One Class SVM, метод изолирующего леса.	6

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	2	Поиск частых наборов с помощью алгоритмов Apriori, ECLAT и FP-Growth.	4
2	2	Поиск шаблонов с помощью мер support, confidence, lift.	4
3	3	Байесовская классификация. Классификация с помощью дерева решений.	4
4	3	Ансамблевая классификация с помощью бэггинга. Ансамблевая классификация с помощью случайного леса. Ансамблевая классификация с помощью бустинга.	4
5	4	Разделительная кластеризация с помощью алгоритмов k-Means и k-Medoids.	2
6	4	Плотностная кластеризация с помощью алгоритма DBSCAN.	2
7	4	Иерархическая кластеризация с помощью различных мер схожести кластеров.	2
8	4	Вычисление качества кластеризации	2
9	5	Поиск точечных аномалий с помощью z-значимости, правила трех сигм, гистограмм.	4
10	5	Поиск коллективных аномалий метода вложенных циклов, метода решеток, кластеризации.	4

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Индивидуальное задание	Tan P.-N., Steinbach M., Karpatne A., Kumar V. Introduction to Data Mining. 2nd Edition. Pearson, 2019. 839 p. Chapter 10. Avoiding False Discoveries, p. 750-808.	2	69,5

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учи-тыва-ется в ПА
1	2	Текущий контроль	Контрольный опрос по теме "Введение в	1	1	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Введение	экзамен

			дисциплину"			в дисциплину". Время на опрос: 15 минут. Макс балл: 1. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)	
2	2	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск частых наборов"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
3	2	Текущий контроль	Практическое задание "Поиск шаблонов. Поиск ассоциативных правил"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
4	2	Текущий контроль	Контрольный опрос по теме "Поиск шаблонов"	1	2	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Поиск шаблонов". Время на опрос: 15 минут. Макс балл: 2. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)	экзамен
5	2	Текущий контроль	Практическое задание "Классификация. Байесовская классификация"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
6	2	Текущий контроль	Практическое задание "Классификация. Деревья решений"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
7	2	Текущий контроль	Практическое задание "Классификация. Бэггинг"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
8	2	Текущий контроль	Практическое задание "Классификация. Случайный лес"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
9	2	Текущий контроль	Практическое задание "Классификация. Бустинг"	1	7	7 баллов: задание полностью выполнено 1-6: задание выполнено частично 0: задание не выполнено	экзамен
10	2	Текущий контроль	Контрольный опрос по теме "Классификация"	1	2	Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Классификация".	экзамен

						<p>Время на опрос: 15 минут. Макс балл: 2.</p> <p>При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)</p>	
11	2	Текущий контроль	Практическое задание "Кластеризация. Разделительная кластеризация"	1	7	<p>7 баллов: задание полностью выполнено</p> <p>1-6: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
12	2	Текущий контроль	Практическое задание "Кластеризация. Плотностная кластеризация"	1	7	<p>7 баллов: задание полностью выполнено</p> <p>1-6: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
13	2	Текущий контроль	Практическое задание "Кластеризация. Иерархическая кластеризация"	1	7	<p>7 баллов: задание полностью выполнено</p> <p>1-6: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
14	2	Текущий контроль	Практическое задание "Кластеризация. Качество кластеризации"	1	7	<p>7 баллов: задание полностью выполнено</p> <p>1-6: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
15	2	Текущий контроль	Контрольный опрос по теме "Кластеризация"	1	2	<p>Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Кластеризация".</p> <p>Время на опрос: 15 минут. Макс балл: 2.</p> <p>При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)</p>	экзамен
16	2	Текущий контроль	Практическое задание "Поиск аномалий. Поиск точечных аномалий"	1	6	<p>6 баллов: задание полностью выполнено</p> <p>1-5: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
17	2	Текущий контроль	Практическое задание "Поиск аномалий. Поиск коллективных аномалий"	1	6	<p>6 баллов: задание полностью выполнено</p> <p>1-5: задание выполнено частично</p> <p>0: задание не выполнено</p>	экзамен
18	2	Текущий контроль	Контрольный опрос по теме "Поиск аномалий"	1	2	<p>Контрольный опрос проводится в виде компьютерного теста по окончании изучения темы "Поиск аномалий".</p> <p>Время на опрос: 15 минут. Макс</p>	экзамен

						балл: 2. При оценивании результатов мероприятия используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)	
19	2	Промежуточная аттестация	Компьютерное тестирование	-	40	<p>Промежуточная аттестация включает компьютерное тестирование. Контрольное мероприятие промежуточной аттестации проводится во время экзамена. Тест состоит из 20 равноценных вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 1 час.</p> <p>На экзамене происходит оценивание учебной деятельности обучающихся по дисциплине на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации.</p> <p>При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)</p> <p>Зачтено: Величина рейтинга обучающегося по дисциплине 60...100 %</p> <p>Не зачтено: Величина рейтинга обучающегося по дисциплине 0...59 %.</p> <p>Допускается выставление оценки на основе текущего рейтинга (автоматом).</p>	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	<p>На экзамене происходит оценивание учебной деятельности обучающихся по дисциплине на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации. При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (утверждена приказом ректора от 24.05.2019 г. № 179)</p> <p>Отлично: Величина рейтинга обучающегося по дисциплине 85...100 % Хорошо: Величина рейтинга обучающегося по дисциплине 75...84 %</p>	<p>В соответствии с пп. 2.5, 2.6 Положения</p>

	Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60...74 % Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0...59 %. Допускается выставление оценки на основе текущего рейтинга (автоматом).	
--	---	--

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
ПК-7	Знает: определение базовых задач машинного обучения (поиск шаблонов, классификация, кластеризация и поиск аномалий) и основные алгоритмы их решения	+			+		+			+					+				+	+
ПК-7	Умеет: определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) *основная литература:*

Не предусмотрена

б) *дополнительная литература:*

Не предусмотрена

в) *отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:*

Не предусмотрены

г) *методические указания для студентов по освоению дисциплины:*

1. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с.

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	eLIBRARY.RU	Алексеев Д.С., Щекочихин О.В. Технологии интеллектуального анализа данных. Учебное пособие. Кострома, 2020. 140 с.

			https://elibrary.ru/item.asp?id=43946965
2	Дополнительная литература	eLIBRARY.RU	Жаров А.Н., Минеичева И.Г. Анализ данных. Ярославль, 2020. 148 с. https://elibrary.ru/item.asp?id=43846458
3	Методические пособия для самостоятельной работы студента	Электронная библиотека Юрайт	Миркин Б.Г. Введение в анализ данных. Учебник и практикум. Москва, 2020. 174 с. https://urait.ru/bcode/432851

Перечень используемого программного обеспечения:

Нет

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Практические занятия и семинары		Персональный компьютер
Лекции		Проектор