

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Новосибирский национальный исследовательский
государственный университет» (Новосибирский государственный университет, НГУ)

Факультет информационных технологий

СОГЛАСОВАНО

Декан ФИТ НГУ

 М.М. Лаврентьев

«03» июля 2019 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Интеллектуальный анализ данных

Направление подготовки: 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Направленность (профиль): Программная инженерия и компьютерные науки

Форма обучения: очная

Год обучения: 4, семестр: 7

№	Вид деятельности	Семестр
		7
1	Лекции, час.	32
2	Практические занятия, час.	32
3	Лабораторные занятия, час.	
4	Занятий в контактной форме без учета промежуточной аттестации, час, из них	64
5	в электронной форме, час.	
6	из них аудиторных занятий, час.	64
7	из них в активной и интерактивной форме, час.	32
8	консультаций, час.	
9	Самостоятельная работа, час.	42
10	в том числе на выполнение письменных работ, час	
11	Форма аттестации (экзамен, зачет, дифференцированный зачет), час	ДЗ, 2
12	Всего зачетных единиц ¹	3

Новосибирск 2019

¹ С учетом выделенных часов на промежуточную аттестацию

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта (ФГОС) высшего образования - бакалавриат по направлению подготовки 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА.

Федеральный государственный образовательный стандарт (ФГОС) высшего образования - бакалавриат по направлению подготовки 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА введен в действие приказом Минобрнауки от 19.09.2017 № 929.

Место дисциплины в структуре учебного плана: Блок 1 Дисциплины (модули); часть, формируемая участниками образовательных отношений, дисциплина по выбору.

Рабочая программа дисциплины утверждена решением Ученого совета факультета информационных технологий от 02.07.2019, протокол № 75.


Программу разработал:

Старший преподаватель кафедры общей информатики ФИТ,
кандидат технических наук



И.А.Борисова


Заведующий кафедрой общей информатики ФИТ,
доктор физико-математических наук



Д.Е.Пальчунов

Ответственный за образовательную программу:

доцент кафедры систем информатики ФИТ,
кандидат технических наук



А.А. Романенко

Аннотация к рабочей программе дисциплины «Интеллектуальный анализ данных»

Дисциплина «Интеллектуальный анализ данных» реализуется в рамках образовательной программы высшего образования – программы бакалавриата 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, направленность (профиль): ПРОГРАММНАЯ ИНЖЕНЕРИЯ И КОМПЬЮТЕРНЫЕ НАУКИ, по очной форме обучения на русском языке.

Место в образовательной программе: Дисциплина «Интеллектуальный анализ данных» развивает знания, умения и навыки, сформированные у обучающихся по результатам изучения следующих дисциплин: «Теория вероятностей и математическая статистика», «Информатика» и «Дискретная математика».

Дисциплина «Интеллектуальный анализ данных» реализуется в 7 семестре в рамках части, формируемой участниками образовательных отношений, дисциплин (модулей) Блока 1 и является дисциплиной по выбору.

Дисциплина «Интеллектуальный анализ данных» направлена на формирование компетенций:

Способен разрабатывать компоненты системных программных продуктов (ПКС-2), в части следующих индикаторов достижения компетенций:

ПКС-2.3 Уметь применять знания в области разработки ПО в предметной области.

ПКС-2.7 Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение.

Перечень основных разделов дисциплины:

Содержание дисциплины охватывает круг вопросов, связанных с задачами, подходами и алгоритмами, возникающими в интеллектуальном анализе данных. Дисциплина «Интеллектуальный анализ данных» имеет своей целью:

– формирование у студентов представления о типах задач, возникающих в области интеллектуального анализа данных (Data Mining).

– изучение основных подходов и алгоритмов решения задач анализа данных и особенностей их применения к решению реальных задач.

– получение студентами навыка по выявлению, формализации и успешному решению практических задач анализа данных, возникающие в процессе их профессиональной деятельности.

– получение практического навыка в работе с существующими программными пакетами по анализу данных.

Уникальность курса заключается в объединении в рамках одного курса как информации об основных мировых тенденциях в области анализа данных, так и новейших результатов работы в данном направлении лаборатории анализа данных Института математики СО РАН, что позволяет использовать полученные навыки и знания при работе над выпускной квалификационной работой бакалавра тем студентам, которые проходят специализацию в этой лаборатории и в лабораториях смежной тематики.

При освоении дисциплины студенты выполняют следующие виды учебной работы: лекции, практические занятия, самостоятельная работа. В учебном процессе предусматривается использование активных и интерактивных форм проведения занятий.

В том числе, предполагаются выступления с докладами и их коллективное обсуждение, обсуждение способов и результатов выполнения заданий, участие в конкурсе на лучшее решение прикладной задачи анализа данных.

Самостоятельная работа включает: выполнение 3х промежуточных заданий на языке Python, подготовку к практическим занятиям по разделам дисциплины, подготовку презентаций докладов, написание рефератов, решение конкурсной задачи, подготовку к экзамену.

Общий объем дисциплины – 3 зачетных единицы (108 часов).

Правила аттестации по дисциплине. Текущий контроль по дисциплине «Интеллектуальный анализ данных» осуществляется по результатам выполнения 3х промежуточных заданий, конкурсного задания, а также по результатам презентации и защиты доклада по тематике дисциплины. По каждому из пунктов выставляется оценка «зачтено» или «не зачтено». Оценка «зачтено» по всем пунктам является одним из условий успешного прохождения промежуточной аттестации.

Оценка "зачтено" за промежуточные и конкурсное задание выставляется при выполнении следующих условий:

- 1) Имеется работающий код по всем заданиям.
- 2) Реализованы все пункты заданий без грубых смысловых ошибок.
- 3) В конкурсном задании результаты распознавания тестовой выборки не хуже датчика случайных чисел;

Оценка "зачтено" за презентацию и доклад выставляется при выполнении следующих условий:

- 1) Презентации и доклады на темы, соответствующие разделам дисциплины выполнены и защищены в полном соответствии с предъявляемыми требованиями (оценка "зачтено");

Промежуточная аттестация по дисциплине «Интеллектуальный анализ данных» проводится по завершению периода ее освоения (семестра).

Промежуточная аттестация по дисциплине проводится в два этапа:

- 1) Оценочное портфолио по результатам работы в семестре, которое включает: 3 промежуточных задания, конкурсное задание, а также презентация и защита доклада по тематике дисциплины..
- 2) Дифференцированный зачет ставится по результатам устного опроса, в процессе которого обучающийся отвечает на три вопроса, согласно вытянутому билету. Во время ответа обучающемуся могут быть заданы дополнительные вопросы, в зависимости от вопросов, образующих билет.

Результаты промежуточной аттестации по дисциплине оцениваются по шкале «неудовлетворительно», «удовлетворительно», «хорошо», «отлично». Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение промежуточной аттестации и ставятся при условии наличия оценки «зачтено» по всем заданиям из портфолио.

Оценка «отлично» соответствует продвинутому уровню сформированности компетенции.

Оценка «хорошо» соответствует базовому уровню сформированности компетенции.

Оценка «удовлетворительно» соответствует пороговому уровню сформированности компетенции.

Учебно-методическое обеспечение дисциплины.

Учебно-методический комплекс по дисциплине «Интеллектуальный анализ данных» в электронной информационно-образовательной среде НГУ:

Материалы по дисциплине «Интеллектуальный анализ данных», включая слайды лекций, тексты примеров программ на языке Python, тексты заданий и сводные результаты по курсу содержатся в облаке: <https://cloud.mail.ru/public/7EDT/i7fskwwjw> .

1. Внешние требования к дисциплине

Таблица 1.1

Компетенция ПКС-2 Способен разрабатывать компоненты системных программных продуктов, в части следующих индикаторов достижения компетенции:	
ПКС-2.3	Уметь применять знания в области разработки ПО в предметной области.
ПКС-2.7	Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение .

2. Требования к результатам освоения дисциплины

Таблица 2.1

Результаты изучения дисциплины по уровням освоения (иметь представление, знать, уметь, владеть)	Формы организации занятий		
	Лекции	Практики / семинары	Самостояте льная работа
ПКС-2.3 Уметь применять знания в области разработки ПО в предметной области.			
1. Знать: структуру и содержимое программных пакетов для анализа данных (например пакет sklearn в языке python)		+	+
2. Уметь: использовать такие пакеты для решения конкретных прикладных задач.		+	+
ПКС-2.7 Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение .			
3. Знать: основные задачи и методы анализ данных	+	+	+
4. Уметь: вычленять задачи анализа данных из практических задач		+	+
5. Уметь выбирать модель данных, наиболее адекватную решаемой задаче, подбирать алгоритм для выбранной модели.	+	+	+
6. Уметь сравнивать результаты решения задачи различными алгоритмами и обосновывать выбор наилучшего алгоритма.	+		+

3. Содержание и структура учебной дисциплины

Таблица 3.1

Темы лекций	Активные формы, час.	Часы	Ссылки на результаты обучения
Семестр: 7			
1. Обзор состояния проблемы искусственного интеллекта		2	3
2. Теория измерений. Методы предварительной обработки данных.		2	3
3. Классификация задач анализа данных. Базовые гипотезы.		2	3,5
4. Задача классификации и основные подходы к ее решению		4	3,5,6
5. Задача кластеризации и основные подходы к ее решению		4	3,5,6
6. Задача частичного обучения и основные подходы к ее решению.		2	3,5,6
7. Задача прогнозирования и основные подходы к ее решению		4	3,5,6
8. Задача выбора и генерации информативной системы		2	3,5,6

признаков			
9. Задача заполнения пробелов в таблицах		2	3,5,6
10. Задача фильтрации шумовых объектов		2	3,5,6
11. Задачи комбинированного типа		2	3,5,6
12. Функция конкурентного сходства и алгоритмы, основанные на ней		2	3
13. Анализ временных рядов		2	3,5,6
Итого		32	

Таблица 3.2

Темы практических занятий	Активные формы, час.	Часы	Ссылки на результаты обучения	Учебная деятельность
Семестр: 7				
Тема 1. Изучение основ языка Python и его стандартных пакетов.	4	4	1,2	Обучающиеся знакомятся с основами языка программирования Python, которые необходимы для дальнейшего его использования в процессе анализа данных.
Тема 2. Изучение пакета sklearn и принципов его использования для решения задач анализа данных	4	4	1,2,3,5	Обучающиеся знакомятся с основными принципами использования пакета sklearn для анализа данных
Тема 3. Использование пакета sklearn для решения прикладных задач анализа данных	10	10	1,2,3,4,5	Обучающиеся учатся использовать пакет для решения модельных и реальных задач анализа данных, знакомятся с принципам подбора алгоритмов и оценки качества их работы
Тема 4. Работа над докладами студентов на темы, связанные с использованием методов анализа данных для решения интересных практических задач	10	10	3,4,5	Обучающиеся выбирают тематику для своих докладов, обсуждают их содержание, презентуют и обсуждают возможности для использования рассматриваемых алгоритмов в своей научной и профессиональной деятельности
Тема 5. Решение теоретических задач, возникающих в процессе анализа данных.	4	4	3,4,5	Обучающимся предлагаются для решения небольшие теоретические задачи, связанные с анализом данных. Сначала эти задачи решаются совместно с преподавателем, а затем самостоятельно.

Итого	32	32		
--------------	-----------	-----------	--	--

4. Самостоятельная работа студентов

Таблица 4.1

№	Виды самостоятельной работы	Ссылки на результаты обучения	Часы на выполнение	Часы на консультации
Семестр: 7				
1	Подготовка к практическим занятиям по теме 1.	1,3	4	0
	Обучающиеся выполняют промежуточное задание 1, текст которого содержится в материалах курса https://cloud.mail.ru/public/7EDT/i7fskwwjw . Это позволяет им познакомиться с языком Python и базовыми принципами его использования, а также на практике освоить теорию измерений и принципы работы с разнотипными описывающими признаками в процессе анализа данных. Результаты выполнения представляются в виде работающей программы .			
2	Подготовка к практическим занятиям по теме 2 и 3.	1,2,3,5,6	18	0
	Обучающиеся выполняют промежуточные задания 2, 3 и конкурсного задания, тексты которых содержится в материалах курса https://cloud.mail.ru/public/7EDT/i7fskwwjw . В этих задачах они учатся решать такие задачи как классификация, выбор информативных признаков, фильтрация шумовых объектов, оценивать качество решения этих задач, подбирать параметры алгоритмов под конкретную задачу, сравнивать качество различных алгоритмов, определять целесообразность тех или иных шагов на этапе предобработки данных в зависимости от того, какими свойствами обладает исходная анализируемая выборка. По результатам выполнения этих заданий, пишется отчет, где анализируются полученные результаты исследований. К отчету прилагаются работающие программы. В конце семестра обучающимся предлагается для решения конкурсная задача, содержание которой содержится в материалах курса https://cloud.mail.ru/public/7EDT/i7fskwwjw . Для ее решения они могут использовать любые методы и алгоритмы, но должны обосновать свой выбор и предоставить результаты прогнозирования целевого признака на тестовой выборке. Это позволяет определить, чей алгоритм оказался объективно самым точным.			
3	Подготовка к практическим занятиям по теме 4.	3,4,5,6	6	0
	Обучающиеся проводят анализ и обоснование выбора темы доклада и после ее согласования собирают и структурируют информацию по ней. Для этого Обучающийся формулирует постановку задачи, описывает алгоритм, позволяющий решать эту задачу, обосновывает успешность этого алгоритма применительно к рассматриваемой задаче, а также описывает спектр прикладных задач, для которых также может быть применен этот алгоритм. По результатам работы оформляется презентация для обсуждения и защиты на практическом занятии.			
4	Подготовка к практическим занятиям по теме 5	3,4,5,6	4	0
	Обучающиеся агрегируют все знания, полученные на теоретических и практических занятиях, чтобы успешно решать предложенные теоретические задачи, связанные с анализом данных.			
5	Подготовка к диф.зачету	1,2,3,4,5,6	10	0
	Подготовка к диф.зачету по вопросам, представленным в фонде оценочных средств, являющихся приложением к рабочей программе дисциплины.			

Итого		42	
--------------	--	-----------	--

5. Образовательные технологии

В ходе реализации учебного процесса по дисциплине проводятся лекционные и семинарские занятия. Практические вопросы, связанные с использованием знаний, полученных на лекциях, рассматриваются на семинарах, и закрепляются в процессе выполнения заданий по курсу, по вопросам, вызывающим затруднения, проводятся консультации.

В ходе реализации учебного процесса по дисциплине применяются такие формы проведения практических занятий, как дискуссии, обсуждение и защита результатов работы.

Для организации и контроля самостоятельной работы студентов, а также проведения консультаций применяются информационно-коммуникационные технологии. (таблица 5.1).

Таблица 5.1

Информирование	https://cloud.mail.ru/public/7EDT/i7fskwwjw
Консультирование	biamia@mail.ru
Контроль	biamia@mail.ru
Размещение учебных материалов	https://cloud.mail.ru/public/7EDT/i7fskwwjw

6. Правила аттестации студентов по учебной дисциплине

По дисциплине «Интеллектуальный анализ данных» проводится текущая и промежуточная аттестация (итоговая по дисциплине).

Текущая аттестация по дисциплине по дисциплине «Интеллектуальный анализ данных» осуществляется по результатам выполнения 3х промежуточных заданий, конкурсного задания, а также по результатам презентации и защиты доклада по тематике, вписывающейся в рамки основных разделов дисциплины. По каждому из пунктов выставляется оценка «зачтено» или «не зачтено». Оценка «зачтено» по всем пунктам является одним из условий успешного прохождения промежуточной аттестации.

Оценка "зачтено" за промежуточные и конкурсное задание выставляется при наличии следующих условий:

- 1) Имеется работающий код на языке Python или другом по выбору обучающегося.
- 2) Все пункты задания реализованы без грубых смысловых ошибок.
- 3) Все стандартные этапы решения задач анализа данных должны быть реализованы корректно. Все получаемые результаты распознавания должны быть не хуже тех, что выдает датчик случайных чисел.
- 4) К заданию должен прилагаться отчет, содержащий описание проводимых экспериментов и обоснованный анализ полученных результатов.

Оценка "зачтено" за презентацию и доклад на тему, соответствующую одному из разделов изучаемой дисциплины по выбору, выставляется при выполнении следующих условий:

- 1) Содержание презентации и доклада соответствует заявленной теме.
- 2) Материал достаточно полно отражает тематику и изложен на понятном аудитории языке.
- 3) Докладчик способен ориентироваться в излагаемом материале и отвечать на дополнительные вопросы преподавателя и других обучающихся.

Промежуточная аттестация (итоговая по дисциплине) проводится по завершению периода ее освоения (семестра). Промежуточная аттестация проводится в форме диф.зачета. Результаты аттестации по дисциплине оцениваются по шкале «неудовлетворительно», «удовлетворительно», «хорошо», «отлично». Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение итоговой аттестации при условии наличия оценки «зачтено» по промежуточным заданиям, презентации и конкурсному заданию.

Итоговая оценка за освоение дисциплины выставляется по результатам оценивания работ студента (портфолио), которые включают:

1. результаты по выполнению 3х промежуточных заданий с работающим программным кодом;
2. презентация и устный доклад на темы, связанные с основными разделами дисциплины;
3. результаты по выполнению конкурсного задания с работающим программным кодом и файлом результатов распознавания тестовой выборки.

2. Диф.зачет

Оценка за дисциплину в 7 семестре выставляется в формате «неудовлетворительно», «удовлетворительно», «хорошо», «отлично». Оценки «отлично», «хорошо», «удовлетворительно» означает успешное освоение дисциплины при условии наличия оценки «зачтено» по промежуточным заданиям, реферату и конкурсному заданию.

Оценка «отлично» соответствует продвинутому уровню сформированности компетенции.

Оценка «хорошо» соответствует базовому уровню сформированности компетенции.

Оценка «удовлетворительно» соответствует пороговому уровню сформированности компетенции.

В таблице 6.1 представлено соответствие форм аттестации заявляемым требованиям к результатам освоения дисциплины.

Таблица 6.1

Коды компетенций ФГОС	Результаты обучения	Формы аттестации	
		Портфолио	Диф.зачет
ПКС.2	ПКС-2.3 Уметь применять знания в области разработки ПО в предметной области.	+	
	ПКС-2.7 Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение .	+	+

Требования к структуре и содержанию портфолио, оценочные средства, а также критерии оценки сформированности компетенций и освоения дисциплины в целом, представлены в Фонде оценочных средств, являющемся приложением 1 к настоящей рабочей программе дисциплины.

7. Литература

1. Загоруйко, Николай Григорьевич. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. Новосибирск : Изд-во Ин-та математики, 1999. 268с. То же [Электронный ресурс]. - URL:http://www.rfbr.ru/rffi/ru/books/o_18614

№ п/п	Наименование Интернет-ресурса	Краткое описание
1	Машинное обучение (курс лекций, К.В.Воронцов) [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%28%D0%BA%D1%83%D1%80%D1%81_%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C_%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE%D0%B2%29 . – Загл. с экрана	Материалы по машинному обучению
2	[Электронный ресурс]. – Режим доступа: http://scikit-learn.org/stable/modules/classes.html – Загл. с экрана	Описание пакета для анализа данных sklearn

8. Учебно-методическое и программное обеспечение дисциплины

8.1. Учебно-методическое обеспечение

Материалы по дисциплине «Интеллектуальный анализ данных», включая слайды лекций, тексты примеров программ на языке Python, тексты заданий и сводные результаты по курсу содержатся в облаке: <https://cloud.mail.ru/public/7EDT/i7fskwwjw>

8.2. Программное обеспечение

Для обеспечения реализации дисциплины используется стандартный комплект программного обеспечения (ПО), включающий регулярно обновляемое лицензионное ПО Windows и MS Office.

Специализированное программное обеспечение не требуется.

9. Профессиональные базы данных и информационные справочные системы

1. Полнотекстовые журналы Springer Journals за 1997-2015 г., электронные книги (2005-2016 гг.), коллекция научных материалов в области физических наук и инжиниринга SpringerMaterials, реферативная БД по чистой и прикладной математике zbMATH.

2. Электронная библиотека диссертаций Российской государственной библиотеки (ЭБД РГБ)

3. Полнотекстовые электронные ресурсы Freedom Collection издательства Elsevier (Нидерланды) (Computer Science;)

4. Электронные ресурсы Web of Science Core Collection (Thomson Reuters Scientific LLC.), Journal Citation Reports + ESI
5. Электронные БД JSTOR (США). Mathematics & Statistics,
6. БД Scopus (Elsevier)
7. Лицензионные материалы на сайте eLibrary.ru

10. Материально-техническое обеспечение

Таблица 10.1

№	Наименование	Назначение
1	Презентационное оборудование (мультимедиа-проектор, экран, компьютер для управления)	Для проведения лекционных занятий
2	Компьютерный класс (с выходом в Internet)	Для организации самостоятельной работы обучающихся

Материально-техническое обеспечение образовательного процесса по дисциплине для обучающихся из числа лиц с ограниченными возможностями здоровья осуществляется согласно «Порядку организации и осуществления образовательной деятельности по образовательным программам для инвалидов и лиц с ограниченными возможностями здоровья в Новосибирском государственном университете».

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Новосибирский национальный исследовательский
государственный университет» (Новосибирский государственный университет, НГУ)

Факультет информационных технологий

СОГЛАСОВАНО



Декан ФИТ НГУ

М.М. Лаврентьев

«03» июля 2019 г.

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
по дисциплине Интеллектуальный анализ данных**

Направление подготовки: 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Направленность (профиль): Программная инженерия и компьютерные науки

Квалификация: бакалавр

Форма обучения: очная

Год обучения: 4, семестр 7

Форма аттестации	Семестр
Дифференцированный зачет	7

Фонд оценочных средств промежуточной аттестации по дисциплине является **Приложением 1** к рабочей программе дисциплины «Интеллектуальный анализ данных», реализуемой в рамках образовательной программы высшего образования – программы бакалавриата 09.03.01 Информатика и вычислительная техника, направленность (профиль): Программная инженерия и компьютерные науки.

Фонд оценочных средств промежуточной аттестации по дисциплине утвержден решением ученого совета факультета информационных технологий, протокол № 75 от 02.07.2019.

Разработчики:

Старший преподаватель кафедры общей информатики ФИТ,
кандидат технических наук



И.А.Борисова

Заведующий кафедрой общей информатики ФИТ,
доктор физико-математических наук



Д.Е.Пальчунов

Ответственный за образовательную программу:

доцент кафедры систем информатики ФИТ,
кандидат технических наук



А.А. Романенко

1. Содержание и порядок проведения промежуточной аттестации по дисциплине

1.1. Общая характеристика содержания промежуточной аттестации

Промежуточная аттестация по дисциплине «Интеллектуальный анализ данных» проводится по завершению периода освоения образовательной программы (семестра) для оценки сформированности компетенций в части следующих индикаторов достижения компетенции (таблица П1.1).

Таблица П1.1

Код	Компетенции, формируемые в рамках дисциплины «Интеллектуальный анализ данных»	Семестр 7	
		Портфолио	Диф.зачет
	ПКС-2 Способен разрабатывать компоненты системных программных продуктов		
ПКС-2.3	Уметь применять знания в области разработки ПО в предметной области.	+	
ПКС-2.7	Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение	+	+

Промежуточная аттестация включает 2 этапа. Часть компетенций оценивается портфолио, в которое входят работы, выполненные в рамках дисциплины. Часть компетенций оценивается диф.зачетом.

Тематика вопросов и заданий к диф. зачету носит комплексный характер, т.к. включает вопросы ситуационно-производственного, практического, а также научно-исследовательского содержания, и включает следующие темы (разделы):

- 1.Эмпирические гипотезы и закономерности
- 2.Теория измерений. Методы предварительной обработки данных.
- 3.Классификация задач анализа данных. Базовые гипотезы.
- 4.Задача классификации и основные подходы к ее решению
- 5.Задача кластеризации и основные подходы к ее решению
- 6.Задача частичного обучения и основные подходы к ее решению.
- 7.Задача прогнозирования и основные подходы к ее решению
- 8.Задача выбора и генерации информативной системы признаков
- 9.Задача заполнения пробелов
- 10.Задача фильтрации шумовых объектов
- 11.Задачи комбинированного типа
- 12.Функция конкурентного сходства и алгоритмы, основанные на ней
- 13.Анализ временных рядов

1.2. Порядок проведения промежуточной аттестации по дисциплине

Промежуточная аттестация проводится в 2 этапа: портфолио и диф. зачет . Необходимым условием для прохождения промежуточной аттестации является оценка «зачтено» по результатам выполненного портфолио. Для оценивания портфолио студенту необходимо сдать все работы, входящие в структуру портфолио – три промежуточных и одно курсное задание, а также презентацию и доклад на тематику курса.

Оценка «зачтено» за выполненные промежуточные и конкурсное задания выставляется при выполнении всех следующих условий:

- 1) При решении каждой из задач обучающийся должен:
 - А) выполнить все подпункты
 - Б) предоставить работающий код,
 - В) предоставить отчет с выводами по результатам проведенных экспериментов,
- 2) Все задачи в задании решены без смысловых ошибок.

Оценка «зачтено» за презентацию и доклад выставляется при выполнении всех следующих условий

- 1) презентация и доклады соответствуют заявленной тематике и достаточно полно отражают основные результаты по ней
- 2) выступающий хорошо ориентируется в теме, отвечает на вопросы участников обсуждения;

Диф. зачет проводится в устной форме. В процессе ответа на вопросы билета к диф. зачету студенту могут быть заданы дополнительные вопросы по темам дисциплины.

2. Требования к структуре и содержанию фонда оценочных средств промежуточной аттестации по дисциплине

Перечень оценочных средств, применяемых на каждом этапе проведения промежуточной аттестации по дисциплине, представлен в таблице П1.2.

Таблица П1.2

№ п/п	Наименование оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде
Семестр 7			
1	Портфолио	Целевая подборка работ студента, раскрывающая его индивидуальные образовательные достижения в одной или нескольких учебных дисциплинах.	Требования к структуре и содержанию портфолио
2	Диф.зачет	Комплекс вопросов	Список теоретических вопросов

2.1 Требования к структуре и содержанию оценочных средств аттестации

2.1.1 Требования к структуре и содержанию портфолио

Портфолио должно содержать:

- 1) Результаты выполнения 3х текущих заданий по темам:
 - основы языка Python
 - решение задач классификации с подбором алгоритмов и их параметров средствами пакета sklearn
 - выбор релевантных признаков и фильтрация шумовых объектов выборки и влияние этих процедур на качество решения задачи классификации
 - 2) Результаты выполнения конкурсного задания по теме:
 - Решение прикладной задачи анализа данных с внешним тестированием качества прогнозов
 - 3) Доклада по тематике:
 - Интересные прикладные задачи и их решения методами анализа данных
- Конкретные темы докладов выбираются студентами и согласуются с преподавателем.

2.1.2 Форма и перечень вопросов для диф.зачета.

Форма билета к диф. зачету

Таблица П1.3

<p>Новосибирский государственный университет Дифференциальный зачет</p>	
<p>Интеллектуальный анализ данных</p> <hr/> <p>наименование дисциплины</p>	
<p>09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА Программная инженерия и компьютерные науки</p> <hr/> <p>наименование образовательной программы</p>	
<p>БИЛЕТ №</p>	
<p>1. Вопрос из категории 1 2. Вопрос из категории 2 3. Вопрос из категории 3</p>	
<p>Составитель</p> <hr/> <p style="text-align: center;">(подпись)</p>	<p style="text-align: right;">И.А. Борисова</p>
<p>Ответственный за образовательную программу</p> <hr/> <p style="text-align: center;">(подпись)</p>	<p>А.А. Романенко</p>
<p>« ____ » _____ 20 ____ г.</p>	

Перечень вопросов к дифф.зачету, структурированный по категориям, представлен в таблице П1.4

Таблица П1.4

Категория	Формулировка вопроса
<p>Категория 1 Базовые понятия (ПКС-2.7)</p>	<ol style="list-style-type: none"> 1. Определение и свойства эмпирических гипотез. 2. Эмпирические закономерности. 3. Теория измерений. Типы измерительных шкал 4. Мера близости между объектами в пространстве сильных признаков. 5. Мера близости между объектами в пространстве порядковых признаков. 6. Меры близости между объектами в пространстве номинальных признаков. 7. Меры близости между сильными признаками 8. Меры близости между порядковыми признаками 9. Меры близости между номинальными признаками 10. Усиление и ослабление признаков. Dumtu-кодирование 11. Гипотеза компактности. Типы компактности.

	12. Функция конкурентного сходства. Измерение компактности на основе конкурентного сходства.
<p>Категория 2 Задачи анализа данных, их классификация и свойства (ПКС-2.7)</p>	<ol style="list-style-type: none"> 1. Классификация задач анализа данных 2. Классификация алгоритмов распознавания 3. Статистическая постановка задачи распознавания. Байесово решающее правило. 4. Оценка качества распознавания. Cross Validation. One Leave Out. 5. Эффект переобученности. Представительность выборки. 6. Задача классификации для случая 3х и более классов. 7. Сравнение качества различных алгоритмов классификации. 8. Формулировка задачи таксономии. Классификация алгоритмов таксономии 9. Виды оптимизируемых критериев качества в задаче таксономии 10. Выбор числа таксонов. 11. Оценка качества таксономии 12. Задача частичного обучения. Основные подходы к ее решению. 13. Задача выбора информативной системы признаков. Классификация алгоритмов выбора признаков 14. Критерии оценки информативности систем признаков 15. Задача построения регрессии(прогнозирования). Классификация методов решения задачи прогнозирования. 16. Задача заполнения пробелов. Метод оценки качества работы алгоритма заполнения пробелов. 17. Задачи комбинированного типа 18. Задача бикластеризации 19. Задача цензурирования обучающей выборки. Классификация подходов к задаче цензурирования. 20. Специфика задачи анализа временных траекторий. Нормализация траекторий методом динамического программирования.
<p>Категория 3 Алгоритмы анализа данных (ПКС -2.7)</p>	<ol style="list-style-type: none"> 1. Байесово решающее правило для случая нормальных распределений. 2. Параметрические и непараметрические алгоритмы восстановления плотностей. 3. Линейный дискриминант Фишера 4. Метод k ближайших соседей и его модификации. 5. Метод потенциальных функций 6. Метод опорных векторов 7. Алгоритмы распознавания, основанные на логических правилах. Деревья решений. 8. Ансамбль классификаторов – способы формирования семейства алгоритмов 9. Ансамбль классификаторов – способы формирования ответа ансамбля 10. Иерархическая таксономия 11. Алгоритмы Forel и Scat. 12. Алгоритм k-means 13. Алгоритм KRAV 14. Алгоритм EM 15. Алгоритм Ad, Del, AdDel, GRAD 16. Поиск в глубину и метод ветвей и границ для выбора признаков 17. Алгоритм СПА 18. Задача генерации информативной признаковой системы. Метод главных компонент

- | |
|---|
| <ol style="list-style-type: none">19. Линейная регрессия.20. Использование регуляризаторов при построении линейной регрессии.21. Непараметрическое ядерное сглаживание в задаче прогнозирования.22. Алгоритм ZET. Его модификации.23. Алгоритм WANGA, его разновидности.24. Задача естественной классификации и алгоритм NatClass25. Алгоритмы цензурирования, основанные на использовании расстояния.26. Цензурирование в многомерных пространствах.27. Алгоритм FRiS-Stolp28. Алгоритм FRiS-Tax29. Алгоритм FRiS-GRAD |
|---|

Набор билетов формируется и утверждается в установленном порядке в начале учебного года при наличии контингента обучающихся, завершающих освоение дисциплины «Интеллектуальный анализ данных» в текущем учебном году.

3. Критерии оценки сформированности компетенций в рамках промежуточной аттестации по дисциплине

Таблица П1.5

Шифр компетенций	Структурные элементы оценочных средств	Показатель сформированности	Не сформирован	Пороговый уровень	Базовый уровень	Продвинутый уровень
ПКС-2	<p>Портфолио</p> <p>Вопрос категории 1 билета к диф. зачету</p> <p>Вопрос категории 2 билета к диф. зачету</p> <p>Вопрос категории 3 билета к диф. зачету</p>	<p>ПКС-2.7</p> <p>Уметь: проводить объективную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение</p>	<p>Демонстрирует отсутствие понимания терминологии предметной области, непонимание основных задач анализа данных и незнание основных алгоритмов их решения</p>	<p>Демонстрирует понимание базовых понятий и задач анализа данных, базовое понимание основных принципов их решения. Знает основные алгоритмы их решения.</p>	<p>Демонстрирует понимание терминологии и базовых понятий, необходимых для анализа данных машинными методами. Демонстрирует знание основных задач анализа данных, умеет их формулировать и опознавать, понимает принципы решения. Демонстрирует знание различных алгоритмов анализа данных и их зоны применимости</p>	<p>Демонстрирует углубленное понимание терминологии и базовых понятий, необходимых для анализа данных машинными методами, умение иллюстрировать их примерами. Демонстрирует знание основных задач анализа данных, умеет их формулировать и опознавать, понимает принципы решения, а также может вычленять свойства задачи, существенные с точки зрения подбора алгоритма ее решения. Демонстрирует знание различных алгоритмов анализа данных, понимание принципов их работы и свойств выбирающих их оптимальное функционирование.</p>

ПКС-2	Портфолио	ПКС-2.3 Уметь применять знания в области разработки ПО в предметной области	Демонстрирует незнание принципов оценки качества алгоритмического решения задач анализа данных	Демонстрирует базовое знание основных принципов оценки качества решения задач анализа данных	Демонстрирует знание основных принципов оценки качества решения различных задач анализа данных и сравнения различных алгоритмов.	Демонстрирует знание основных методов оценки качества решения различных задач анализа данных и сравнения различных алгоритмов. Умеет выбрать лучший алгоритм по совокупности характеристик.
-------	-----------	--	--	--	--	---

4. Правила принятия решения об уровне сформированности компетенций по результатам промежуточной аттестации по дисциплине

В случае, если сформированность компетенции проверяется на различных этапах испытаний, то комплексная оценка по результатам нескольких испытаний о пороговом – базовом – продвинутом уровне выставляется по результатам последнего испытания.

Таблица П1.6

Код	Компетенции, формируемые в рамках дисциплины «Интеллектуальный анализ данных»	7 семестр		Основание для принятия решения о сформированности компетенций
		Портфолио	Экзамен	
ПКС-2.3	Уметь применять знания в области разработки ПО в предметной области.	+		Оценка по результатам портфолио
ПКС-2.7	Уметь: проводить объектную декомпозицию информационной системы, вырабатывать и обосновывать архитектурное решение	+	+	Комплексная оценка по результатам всех этапов

5. Критерии выставления оценок по результатам промежуточной аттестации по дисциплине

Результаты промежуточной аттестации в 7 семестре выставляются по результатам диф. зачета и определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Оценки «отлично», «хорошо», «удовлетворительно» означают успешное прохождение промежуточной аттестации.

Оценка «отлично» выставляется, если обучающийся демонстрирует знание терминологии предметной области, знание всех рассматриваемых задач анализа данных, принципов их решения, умеет вычленять задачи в практических приложениях. Также демонстрирует знание основных алгоритмов решения этих задач и принципов их работы, знает, как оценивать качество решения конкретной задачи, конкретного алгоритма и уметь подбирать алгоритмы под особенности решаемой задачи. Подтверждает эти знания в процессе решения прикладных задач.

Оценка «хорошо» выставляется, если обучающийся демонстрирует знание терминологии предметной области, знание всех рассматриваемых задач анализа данных, принципов их решения, основных алгоритмов анализа данных, а также знает, как оценивать качество решения конкретной задачи, конкретного алгоритма и подтверждает эти знания в процессе решения прикладных задач.

Оценка «удовлетворительно» выставляется, если обучающийся демонстрирует знание терминологии предметной области, знание основных рассматриваемых задач анализа данных, несколько базовых алгоритмов анализа данных, а также знает, как оценивать качество решения конкретной задачи и подтверждает эти знания в процессе решения прикладных задач.

Оценка «неудовлетворительно» выставляется, если обучающийся демонстрирует незнание терминологии предметной области, незнание основных рассматриваемых задач анализа данных и алгоритмов их решения. Не умеет оценивать качество решения конкретной задачи.

Итоговая оценка результатов промежуточной аттестации выставляется по ее результатам при условии оценки «зачтено» по результатам портфолио.

