МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

МЕТОДОЛОГИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Методология интеллектуального анализа данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика (утв. приказом Минобрнауки России от 19.09.2017г. №916) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Шахгельдян К.И.

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 27.05.2026 , протокол № 5

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1582918206
Номер транзакции	0000000000F6FEF3
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью дисциплины является освоение методологии анализа данных и получение навыков проведения комплексного интеллектуального анализа от начала до подготовки аналитического отчета

Задачи освоения дисциплины:

1. Познакомиться с методологией проведение комплексного интеллектуального анализа данных.
2. Систематизировать свои знания по методам сбора, предварительной обработки и анализу данных.
3. Получить навыки комплексного анализа данных, эффективно применяя методологию интеллектуального анализа и полученные ранее знания в дисциплинах магистерской программы.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-3 : Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД1	Знание	Этапов анализа данных
			РД2	Умение	Поставить задачу анализа данных
			РД3	Навык	Составление аналитических отчетов
		ОПК-3.2к : Формулирует обоснованные выводы и рекомендации по предлагаемым техническим и программным решениям по профилю деятельности	РД4	Умение	Находить, собирать и обрабатывать данные
			РД5	Знание	Методов статистического анализа и машинного обучения
			РД6	Навык	использования инструментальных средств интеллектуального анализа данных
	ОПК-4 : Способен применять на практике новые научные принципы и методы исследований	ОПК-4.1к : Выбирает теоретико-методологическую основу инновационных научных исследований	РД1	Знание	Этапов анализа данных
			РД2	Умение	Поставить задачу анализа данных
			РД3	Навык	Составление аналитических отчетов

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	Формирование ценностей	Целевые ориентиры
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Гуманизм	Системное мышление
Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Взаимопомощь и взаимоуважение	Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

Дисциплина относится обязательной части учебного плана, блок 1 дисциплины (модули)

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость	Объем контактной работы (час)						СРС	Форма аттестации
				(З.Е.)	Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА	КСР		
09.04.03 Прикладная информатика	ОФО	М01.Б	3	4	41	8	32	0	1	0	103	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Методология анализа данных	РД1	1	4	0	10	Проект (часть по теме 1)

2	Постановка задачи, определение источников и сбор данных	РД2, РД4	2	6	0	10	Проект (часть по теме 2)
3	Разведочный анализ и предварительная обработка данных	РД6	2	8	0	11	Проект (часть по теме 3)
4	Методы анализа данных	РД3, РД5, РД6	3	14	0	80	Проект (аналитический раздел)
Итого по таблице			**8**	**32**	**0**	**111**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Методология анализа данных.

Содержание темы: Этапы анализа. Постановка задачи, источники и сбор данных, корректировка задачи. Разведочный анализ данных и предварительная обработка данных. Анализ данных. Интерпретация полученных результатов. Разработка аналитических отчетов.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 2 Постановка задачи, определение источников и сбор данных.

Содержание темы: Предметная область и проблема постановки задачи. Методы декомпозиции и постановка задачи анализа данных. Выбор источников данных, статистические и первичные данные. Публичные данные. Корпоративные данные. Оценка данных и корректировка поставленной аналитической задачи.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция и практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 3 Разведочный анализ и предварительная обработка данных.

Содержание темы: Выявление основных структур, выявление проблем данных, выбросы, отклонения. Использование диаграмм и графиков для разведочного анализа. Решение проблем с данными, заполнение пропущенных данных. Проверка гипотез (сравнение групп, проверка на соответствия распределениям). .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция и практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

Тема 4 Методы анализа данных.

Содержание темы: Выбор методы в соответствие с задачей. Особенности методов статистического анализа, анализ временных рядов, использование методов машинного обучения и нейронных сетей. Метрики моделей. Построение моделей, их валидация, объединение. Кросс-валидация. Построение прогнозов. Требования к аналитическим отчетам.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция и практические занятия.

Виды самостоятельной подготовки студентов по теме: выполнение практического задания.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

В ходе изучения дисциплины студенты должны посещать аудиторные занятия. Особенность изучения дисциплины состоит в освоении системного подхода к анализу данных: от постановки задачи, данных, к методам их обработки и анализа, к аналитическому отчету, включающими результаты и выводы. Особое место в овладении частью тем данной дисциплины отводится самостоятельной работе, при этом во время аудиторных занятий могут быть рассмотрены и проработаны наиболее важные и трудные вопросы по той или иной теме дисциплины, а процедуры дообучения моделей, их полная валидация вынесены на самостоятельное обучение.

В соответствии с учебным планом направления подготовки процесс изучения дисциплины предусматривает проведение лекций, практических занятий, консультаций, а также самостоятельную работу студентов.

Ниже перечислены предназначенные для самостоятельного изучения студентами те вопросы, которые во время проведения аудиторных занятий изучаются недостаточно или изучение которых носит обзорный характер.

1. Новые методы снижения размерности пространства признаков (UMAP)
2. Метод кластеризации (DBSCAN)
3. Метод кластеризации t-SNE
4. стохастический градиентный бустинг
5. Анализ выживаемости

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и

навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/560311 (дата обращения: 14.01.2025).

2. Гасанов, Э. Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для вузов / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2026. — 271 с. — (Высшее образование). — ISBN 978-5-534-08684-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/584574 (дата обращения: 19.05.2026).

3. Миркин, Б. Г. Базовые методы анализа данных : учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/560414 (дата обращения: 01.09.2025).

7.2 Дополнительная литература

1. Демидова, Л. А. Разведочный анализ данных. Python : учебно-методическое пособие / Л. А. Демидова. — Москва : РТУ МИРЭА, 2023 — Часть 2— 2023. — 92 с. — ISBN 978-5-7339-1933-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/382691 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

2. Математические и программные методы построения моделей глубокого обучения : учебное пособие / А. В. Протодьяконов, А. В. Дягилева, П. А. Пылов, Р. В. Майтак. - Москва ; Вологда : Инфра-Инженерия, 2023. - 176 с. - ISBN 978-5-9729-1484-5. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2094440 (дата обращения: 31.05.2026)

3. Пальмов, С. В. Интеллектуальные информационные системы и технологии : методические указания / С. В. Пальмов. — Самара : ПГУТИ, 2024. — 120 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/463613 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Образовательная платформа "ЮРАЙТ"
2. Электронно-библиотечная система "ZNANIUM.COM"
3. Электронно-библиотечная система "ЛАНЬ"
4. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
5. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/

6. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Ист.бесп.эл.питания Smart-UPS 3000VA
- Мульт проектор №1 Panasonic PT-LX26HE
- Облачный монитор 23" LG CAV42K
- Сетевой монитор:Нулевой клиент Samsung SyncMaster NC240
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Microsoft Windows SL 8.1 Russian
- Python

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

МЕТОДОЛОГИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-3 : Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем
		ОПК-3.2к : Формулирует обоснованные выводы и рекомендации по предлагаемым техническим и программным решениям по профилю деятельности
	ОПК-4 : Способен применять на практике новые научные принципы и методы исследований	ОПК-4.1к : Выбирает теоретико-методологическую основу инновационных научных исследований

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Компетенция ОПК-3 «Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД 1	Знание	Этапов анализа данных	корректная последовательность проведение анализа данных
	РД 2	Умение	Поставить задачу анализа данных	сформированный навык постановки задачи
	РД 3	Навык	Составление аналитических отчетов	сформированный навык подготовки аналитических отчетов
ОПК-3.2к : Формулирует обоснованные выводы и рекомендации по предлагаемым техническим и программным решениям по профилю деятельности	РД 4	Умение	Находить, собирать и обрабатывать данные	сформированный навык сбора и обработки данных
	РД 5	Знание	Методов статистического анализа и машинного обучения	способность использоваться методы анализа данных (статистические и машинного обучения)
	РД 6	Навык	использования инструментальных средств интеллектуального анализа данных	сформировавшийся навык использования инструментальн

				ых средств интеллектуальног о анализа данных

Компетенция ОПК-4 «Способен применять на практике новые научные принципы и методы исследований»

Таблица 2.2 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-4.1к : Выбирает теоретико-методологическую основу инновационных научных исследований	РД 1	Знание	Этапов анализа данных	корректная последовательность проведение анализа данных
	РД 2	Умение	Поставить задачу анализа данных	сформированный навык постановки задачи
	РД 3	Навык	Составление аналитических отчетов	сформированный навык подготовки аналитических отчетов

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очная форма обучения				
РД1	Знание : Этапов анализа данных	1.1. Методология анализа данных	Проект	не предусмотрен
РД2	Умение : Поставить задачу анализа данных	1.2. Постановка задачи, определение источников и сбор данных	Проект	не предусмотрен
РД3	Навык : Составление аналитических отчетов	1.4. Методы анализа данных	не предусмотрен	Проект
РД4	Умение : Находить, собирать и обрабатывать данные	1.2. Постановка задачи, определение источников и сбор данных	Проект	не предусмотрен
РД5	Знание : Методов статистического анализа и машинного обучения	1.4. Методы анализа данных	не предусмотрен	Проект
РД6	Навык : использования инструментальных сред	1.3. Разведочный анализ и предварительная обработка данных	Проект	Проект

	ств интеллектуального анализа данных	1.4. Методы анализа данных	Проект	Проект

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Промежуточная аттестация по дисциплине «Методология анализа данных» включает в себя выполнение практических (составляющих единое проектное задание) и самостоятельных работ, а также обязательную подготовку и сдачу итогового аналитического отчета. Задания, выполенные в рамках проекта позволия оценить уровень усвоения обучающимися знаний, умений и навыков и выявить степень сформированности компетенций. В конце курса проводится экзамен (в виде защиты итогового отчета).

Объем и качество освоения обучающимися дисциплины, уровень сформированности дисциплинарных компетенций оцениваются по результатам текущих и промежуточной аттестаций количественной оценкой, выраженной в баллах, максимальная сумма баллов по дисциплине равна 100 баллам.

Практические задания (части 1,2, 3 проекта) оцениваются по 15 баллов каждая, являются (текущая аттестация).

Практическое задание (итоговое задание в рамках проекта) по промежуточной аттестации оценивается в 40 баллов.

Итоговая аттестация проводится в форме защиты проектного задания, оценивается до 15 баллов.

Сумма баллов, набранных студентом по дисциплине, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика уровня освоения дисциплины
от 91 до 100	«отлично»	Студент демонстрирует сформированность дисциплинарной компетенции на итоговом уровне, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«хорошо»	Студент демонстрирует сформированность дисциплинарной компетенции на среднем уровне: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций на базовом уровне: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по дисциплинарной компетенции, студент испытывает значи

		тельные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций на уровне ниже базового, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не удовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Проект

Выбрать источники данных в соответствии с поставленной задачей, разработать механизм сбора данных для дальнейшего анализа. Проанализировать данные, выявить проблемы, используя графики, диаграммы. Оценить данные, их ограничения для поставленной задачи. Скорректировать и уточнить постановку анализа данных.

Краткие методические указания

Студенты должны использовать поиск в сети Интернет для подбора источника данных, оценивать актуальность и валидность источника. Для сбора данных и оцифровки использовать скрипты из свободного доступа или разработать самостоятельно. Использовать язык R для предварительной обработки данных. Подготовить отчет о результатах выполнение практического задания.

Шкала оценки

Маскимально возможный балл за практическое задание составляет 15 баллов. Для его получения должен быть собран датасет, готовый к дальнейшей обработке в формате xls, а также подготовлен отчет с описанием источника данных, методов сбора и формализации, результатах предварительной обработки.

От 9 до 14 баллов будут получены в случае, если отчет содержит неполную информацию о проделанной работе

от 5 до 8 баллов студент может получить в случае сбора датасета, без описания методов и результатов предварительной обработки.

5.2 Проект

Реализовать предварительную обработку данных, использовать различные схемы корректировки данных и оценить их валидность для поставленной задачи. Выполнить статистический анализ данных методами дескриптивной статистики.

Краткие методические указания

Для выполнения задания студенты должны активно использовать методы оценки качества данных, коррективровки, создания дамми-переменных. Основной язык анализа R. Отчет должен содержать этапы работы, результаты обработки в табличной, текстовой и графическо форме. Использование сложных графиков и диаграмм является желательным.

Шкала оценки

Маскимально возможный балл за практическое задание составляет 15 баллов. Для его получения должен быть проанализирован датасет, выделены все проблемные данные, предложены варианты их корректировки с оценкой их ограничений, а также подготовлен отчет с описанием этапов работы и результатов обработки.

От 9 до 14 баллов будут получены в случае, если отчет содержит неполную информацию о проделанной работе

от 5 до 8 баллов студент может получить в случае поиска и коррективроки основных проблем с данными, без описания методов и результатов предварительной обработки.

5.3 Проект

Проанализировать данные методами статистического анализа. Проверить гипотезы о сходстве и различии в группах. Выявить возможные предикторы. Используя методы снижения размерности свернуть пространство признаков и выявить наиболее значимые признаки. Подготовить аналитический отчет

Краткие методические указания

Использовать методы статистического анализа: проверки гипотез, оценку шансов, корреляционных анализ, дисперсионный анализ, а также методы снижения размерности (PCA, FA, UMAP). Для реаоизации обработки и анализа использовать инструменты библиотек Python (или R).

Шкала оценки

Маскимально возможный балл за практическое задание составляет 15 баллов. Для его получения должен быть проанализирован датасет, выделены и обоснованы потенциальные предикторы, обоснованы методы анализа, а также подготовлен отчет с описанием этапов работы и результатов обработки.

От 9 до 14 баллов будут получены в случае, если отчет содержит неполную информацию о проделанной работе

от 5 до 8 баллов студент может получить в случае выбора предикторов с помощью только проверки гипотез, оценки шансов и корреляционного анализа.

5.4 Проект

Выполнить анализ датасета для выявления и обоснования предикторов, построить модели, сравнить их точность, используя различные метрики качества. Используя методы кластеризации разделить данные на группы, выявить особенности групп. Построить модели, используя различные методы машинного обучения: линейную, логистическую, мультиномиальную, порядковую регрессии, случайный лес, метод опорных векторов, стохастический градиентный бустинг. Построить модели, используя нейронные сети. Используя метрики точности, сравнить модели.

Подготовить аналитический отчет по выполненному анализу, оформить его, используя графики, диаграммы, таблицы. Показать характерные особенности, представить результат прогнозирования и сравнения различные моделей.

Краткие методические указания

Использовать для поиска предикторов методы отбора предикторов. При построении модели сравнивать модели между собой с различным набором предикторов. Обосновывавать выбор метрик. Для валидации моделей использовать процедуру кросс-валидации.

Результаты исследования должны быть оформлены в аналитический отчет с обоснованием полученных результатов.

Шкала оценки

Маскимально возможный балл за практическое задание составляет 40 баллов. Для его получения должен быть проанализирован датасет, выделены и обоснованы предикторы, выполненно сравнение моделей, полученных различными методами, а также подготовлен отчет с описанием этапов работы и результатов обработки.

От 35 до 40 баллов будут получены в случае, если отчет содержит неполную информацию о проделанной работе

от 20 до 34 баллов студент может получить в случае выбора предикторов с помощью только регрессионных методов и одного-двух дургих методов. Отчет содержит неполную информацию о проделанной работе.

от 10 до 19 баллов студент может получить в случае выбора предикторов с помощью только регрессионных методов с кратким описанием исследования.

Ключи для ФОС

5.1 Для проектного задания по дисциплине «Методология интеллектуального анализа данных»

Критерий	**Описание**
Подбор источников данных, оценка актуальности и валидности источников наборов данных	Выбор источников данных в соответствии с поставленной задачей. Выбор (или разработка) «механизма» сбора данных для дальнейшего анализа
Анализ качества наборов данных для интеллектуальной модели	Выбор и обоснование выбора методов / тестов / для анализа качества и характера данных располагаемого набора данных
Интерпретация результатов анализа качества и характера данных исходного датафрейма	Аргументированные выводы о качестве данных: необходимости очистки, предварительной обработки, дополнении, генерации данных
Предварительная обработка данных, формирование наборов данных пригодных для применения методов машинного обучения	Использование методов оценки качества данных, корректировки данных, создания дамми-переменных. Использование сложных графиков и диаграмм.
Анализ данных методами прикладной статистики	Проверка гипотез о сходстве и различии в группах. Выявление возможных предикторов. Снижение размерности пространства признаков, выбор наиболее значимых признаков.
Выбор методов, моделей МО соответствующих задаче	Анализ датасета для выявления и обоснования предикторов. Построенные модели. Сравнительный анализ их точности. Обоснование используемых метрик качества.
Интерпретация	Оценка эффективности построенной интеллектуальной модели, оценка прогностической силы построенной модели МО
Использование инструментария	Реализация с использование python (или Rstudio)
Достигнутые результаты	Умение студента реализовать полный цикл интеллектуальной обработки данных соответствующих практической задаче, типу данных