

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Рабочая программа дисциплины (модуля)
МЕТОДОЛОГИЯ АНАЛИЗА ДАННЫХ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Искусственный интеллект и машинное обучение в
управлении и принятии решений

Год набора на ОПОП
2022

Форма обучения
очная

Владивосток 2023

Рабочая программа дисциплины (модуля) «Методология 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., доктор технических наук, профессор, Кафедра информационных технологий и систем, carina.shahgeldyan@vvsu.ru

Утверждена на заседании кафедры информационных технологий и систем от 31.05.2023 , протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	0000000000BBE2E8
Владелец	Кийкова Е.В.

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 проведения 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	РД4	Умение	Находить, собирать и обрабатывать данные
	ОПК-6 : Способен исследовать современные проблемы и методы прикладной информатики и развития информационного общества	ОПК-6.2к : Формирует представление об оценке качества информации в информационных системах	РД5	Знание	Методов статистического анализа и машинного обучения
	ОПК-7 : Способен использовать методы научных исследований и математического моделирования в области проектирования и управления	ОПК-7.2к : Моделирует управленческие решения для различных процессов и проводит их сравнительный анализ для	РД3	Навык	Составление аналитических отчетов

	информационными системами	принятия решений	РД5	Знание	Методов статистического анализа и машинного обучения
	УК-1 : Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.2к : Принимает конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий	РД1	Знание	Этапов анализа данных
			РД2	Умение	Поставить задачу анализа данных

2 Место дисциплины (модуля) в структуре ОПОП

Входными требованиями к прохождению курса является освоенные ранее дисциплины по методам анализа и технологиям сбора данных.

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)	Трудо-емкость (З.Е.)	Объем контактной работы (час)					СРС	Форма аттес-тации	
					Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА			КСР
09.04.03 Прикладная информатика	ОФО	М01.Б	3	4	33	8	24	0	1	0	111	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 ре-зультата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Методология анализа данных	РД1	1	4	0	10	защита практического задания

2	Постановка задачи, определение источников и сбор данных	РД2, РД4	2	4	0	10	защита практического задания
3	Разведочный анализ и предварительная обработка данных		2	4	0	11	защита практического задания
4	Методы анализа данных	РД3, РД5	3	12	0	80	защита практического задания
Итого по таблице			8	24	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 соответствии 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 аудиторские 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)

0.1 Основная литература

0.2 Дополнительная литература

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

Отсутствуют

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

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

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

МЕТОДОЛОГИЯ АНАЛИЗА ДАННЫХ

Направление и направленность (профиль)

09.04.03 Прикладная информатика. Искусственный интеллект и машинное обучение в
управлении и принятии решений

Год набора на ОПОП
2022

Форма обучения
очная

Владивосток 2023

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-6 : Способен исследовать современные проблемы и методы прикладной информатики и развития информационного общества	ОПК-6.2к : Формирует представление об оценке качества информации в информационных системах
	ОПК-7 : Способен использовать методы научных исследований и математического моделирования в области проектирования и управления информационными системами	ОПК-7.2к : Моделирует управленческие решения для различных процессов и проводит их сравнительный анализ для принятия решений
	УК-1 : Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.2к : Принимает конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий
УК-1.2к : Принимает конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий		

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 результата	Тип результата	Результат	
ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД4	Умение	Находить, собирать и обрабатывать данные	сформированный навык сбора и обработки данных

Компетенция ОПК-6 «Способен исследовать современные проблемы и методы прикладной информатики и развития информационного общества»

Таблица 2.2 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код результата	Тип результата	Результат	
ОПК-6.2к : Формирует представление об оценке качества информации в информационных системах	РД5	Знание	Методов статистического анализа и машинного обучения	способность использовать методы анализа данных (статистические и машинного обучения)

Компетенция ОПК-7 «Способен использовать методы научных исследований и математического моделирования в области проектирования и управления информационными системами»

Таблица 2.3 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код результата	Тип результата	Результат	
ОПК-7.2к : Моделирует управленческие решения для различных процессов и проводит их сравнительный анализ для принятия решений	РД3	Навык	Составление аналитических отчетов	сформированный навык подготовки аналитических отчетов
	РД5	Знание	Методов статистического анализа и машинного обучения	способность использовать методы анализа данных (статистические и машинного обучения)

Компетенция УК-1 «Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий»

Таблица 2.4 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код результата	Тип результата	Результат	

УК-1.2к : Принимает конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий	РД1	Знание	Этапов анализа данных	корректная последовательность проведения анализа данных
	РД2	Умение	Поставить задачу анализа данных	сформированный навык постановки задачи

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	Практическая работа	Контрольная работа
РД2	Умение : Поставить задачу анализа данных	1.2. Постановка задачи, определение источников и сбор данных	Практическая работа	Контрольная работа
РД3	Навык : Составление аналитических отчетов	1.4. Методы анализа данных	Практическая работа	Практическая работа
РД4	Умение : Находить, собирать и обрабатывать данные	1.2. Постановка задачи, определение источников и сбор данных	Практическая работа	Контрольная работа
РД5	Знание : Методов статистического анализа и машинного обучения	1.4. Методы анализа данных	Практическая работа	Практическая работа

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 и самостоятельных работ, а также обязательную подготовку и сдачу итогового аналитического отчета. Перечень предложенных работ позволит 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 работы оцениваются по 15 баллов каждая из трех работ текущей аттестации.

Практическая работа (итоговая практическая работа) по промежуточной аттестации оценивается в 50 баллов.

Экзамен проводится в форме защиты итоговой практической работы, оценивается до 5 баллов.

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 и 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 Примеры заданий для выполнения практических работ

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 выполнения практического задания.

Шкала оценки

Максимально 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 Отчет

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, корректировки, создания дамми-переменных. Основным языком анализа R. Отчет должен содержать этапы работы, результаты обработки в табличной, текстовой и графической форме. Использование сложных графиков и диаграмм является желательным.

Шкала оценки

Максимально 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 корректировки основных проблем с данными, без описания методов и результатов предварительной обработки.

5.3 Отчет

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, корреляционный анализ, дисперсионный анализ, а также методы снижения размерности (PCA, FA, UMAP). Для реорганизации обработки и анализа использовать инструменты библиотек Python (или R).

Шкала оценки

Максимально 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 Экзаменационная Контрольная работа

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. Обосновывать выбор метрик. Для валидации моделей использовать процедуру кросс-валидации.

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

Шкала оценки

Максимально возможный балл за практическое задание составляет 50 баллов. Для его получения должен быть проанализирован датасет, выделены и обоснованы предикторы, выполнено сравнение моделей, полученных различными методами, а также подготовлен отчет с описанием этапов работы и результатов обработки.

От 35 до 49 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 других 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.