МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Направление и направленность (профиль)
38.04.05 Бизнес-информатика. Информационная бизнес-аналитика

Год набора на ОПОП
2026

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Методы машинного обучения» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 38.04.05 Бизнес-информатика (утв. приказом Минобрнауки России от 12.08.2020г. №990) 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):

Костерин В.В.

Кригер А.Б.

Назаров Д.А.

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	0000000000F73C17
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Цели:

- изучение теоретических основ и алгоритмов машинного обучения;
- формирование умений и навыков автоматизации решения профессиональных задач в различных областях человеческой деятельности на основе интеллектуального анализа
- данных.

Задачи:

- расширение и углубление теоретических знаний о качественных особенностях
- различных систем, количественных взаимосвязях и закономерностях их развития;
- овладение методологией построения, анализа и использования регрессионных моделей, как для анализа состояния, так и для прогнозных оценок развития изучаемых систем;
- изучение алгоритмов кластеризации и классификации объектов;
- получение навыков реализации методов машинного обучения с учителем и без учителя на языке Python

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	
38.04.05 «Бизнес-информатика» (М-БИ)	ОПК-3 : Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем	ОПК-3.1к : Использует системы искусственного интеллекта в рамках стратегического планирования и прогнозирования в профессиональной деятельности	РД1	Знание	аналитических методов обработки статистических данных
			РД2	Умение	применять методы машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
			РД3	Навык	анализа и оценки результатов МО
		ОПК-3.2к : Использует методы, модели, алгоритмы, технологии и инструментальные средства для работы с	РД4	Знание	методов исследования современных проблем и методов прикладной информатики
			РД5	Умение	разрабатывать оригинальные алгоритмы с использованием языка программирования Python

	искусственного интеллекта	большими данными для поддержки стратегических управленческих решений			для решения профессиональных задач
			РД6	Навык	разработки приложений на языке Python, реализующие методы МО
	ПКВ-2 : Способен выбирать и осуществлять рациональные решения по ИС и ИКТ для управления организациями на основе больших данных	ПКВ-2.1к : Использует и совершенствует методы, модели, алгоритмы, технологии и инструментальные средства для работы с большими данными	РД4	Знание	методов исследования современных проблем и методов прикладной информатики
			РД5	Умение	разрабатывать оригинальные алгоритмы с использованием языка программирования Python для решения профессиональных задач
			РД6	Навык	разработки приложений на языке Python, реализующие методы МО
		ПКВ-2.2к : Разрабатывает технические проекты в сфере ИТ, позволяющие анализировать и использовать большие данные в управлении организацией	РД1	Знание	аналитических методов обработки статистических данных
			РД2	Умение	применять методы машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
			РД3	Навык	анализа и оценки результатов МО
			РД4	Знание	методов исследования современных проблем и методов прикладной информатики

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 к обязательной части Блока 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 обуче-ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо-емкость	Объем контактной работы (час)						СРС	Форма аттес-тации
				(З.Е.)	Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА	КСР		
38.04.05 Бизнес-информатика	ОФО	М01.Б	1	4	55	18	36	0	1	0	89	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	Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python	РД1, РД2, РД3, РД5	4	8	0	25	практическая работа № 1
2	Множественная регрессия и корреляция	РД1, РД2, РД3, РД5, РД6	2	6	0	25	практическая работа № 2
3	Снижение размерности. Метод главных компонент.	РД1, РД2, РД3, РД4, РД5	2	7	0	25	практическая работа № 3
4	Задачи классификации. Модели с дискретной зависимой переменной	РД1, РД2, РД3, РД5, РД6	4	7	0	25	практическая работа № 4
5	Кластерный анализ	РД2, РД3, РД4, РД5, РД6	4	8	0	25	практическая работа № 5
Итого по таблице			**16**	**36**	**0**	**125**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python.

Содержание темы: Общая постановка задачи машинного обучения. Способы машинного обучения.Классические задачи машинного обучения. Регрессия. Спецификация модели, связанная с выбором факторов и вида математической функции. Методы выбора математической функции в парной регрессии. Линейная регрессия. Оценка параметров. Оценка значимости уравнения и параметров. Нелинейная регрессия. Ошибки аппроксимации.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: стандартная, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, подготовка к итоговому тесту.

Тема 2 Множественная регрессия и корреляция.

Содержание темы: Методы построения моделей и спецификация ее систематической части. Мультиколлинеарность. Оценка параметров моделей и их диагностика. Качественные объясняющие переменные. Анализ вклада отдельных предикторов и селекция их оптимальной комбинации. Оценка качества каждой модели-претендента. Предпосылки МНК.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: стандартная, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, подготовка к итоговому тесту.

Тема 3 Снижение размерности. Метод главных компонент.

Содержание темы: Отбор и выделение признаков. Математическая модель метода главных компонент. Оценка числа главных компонент. Оценка уровня информативности и подбор названий для главных компонент.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: стандартная, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, подготовка к итоговому тесту.

Тема 4 Задачи классификации. Модели с дискретной зависимой переменной.

Содержание темы: Модели бинарного выбора. Логит-модель, пробит-модель. Оценивание параметров модели. Метод максимального правдоподобия. Модели множественного выбора с неупорядоченными альтернативами. Модели множественного выбора с упорядоченными альтернативами.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: стандартная, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, подготовка к итоговому тесту.

Тема 5 Кластерный анализ.

Содержание темы: Формальное описание задачи кластеризации. Математические характеристики кластера. Меры близости: меры расстояния, меры сходства, меры предсказуемости вероятностные коэффициенты сходства. Алгоритмы кластерного анализа. Иерархическая кластеризация. Алгоритмы неиерархического разделения. Метод нечетких к-средних. Оценка качества кластеризации.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: стандартная, метод кооперативного обучения.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, подготовка к итоговому тесту.

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

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

На первом шаге выполнения практических работ необходимо сформировать массив данных из указанных источников, согласно рекомендациям и/или загрузить имеющиеся данные в Python.

Далее следует выполнить задания к практической работе, сделать необходимые выводы и сформировать требуемые отчеты.

При подготовке к текущей аудиторной работе рекомендуется использовать материалы лекций, источники из списка учебной литературы, разобрать примеры скриптов, документацию библиотек (пакетов), проанализировать приемы работы с Python.

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.

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. Васильев, Е. П. Интеллектуальный анализ данных в технологиях принятия решений : учебное пособие / Е. П. Васильев, В. И. Орешков. — Рязань : РГРТУ, 2023. — 180 с. — ISBN 978-5-7722-0344-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/380378 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

2. Гафаров, Ф. М. Нейронные сети в PyTorch : учебное пособие / Ф. М. Гафаров, А. Ф. Гилемзянов. - Казань : Казанский федеральный университет, 2024. - 106 с. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2173433 (дата обращения: 31.05.2026)

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, 2024. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/556941 (дата обращения: 12.03.2025).

7.2 Дополнительная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2024. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/557384 (дата обращения: 12.11.2024).

2. Гуриков, С. Р. Основы алгоритмизации и программирования на Python : учебное пособие / С.Р. Гуриков. — Москва : ИНФРА-М, 2025. — 343 с. — (Высшее образование). - ISBN 978-5-16-020255-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2166199 (дата обращения: 31.05.2026)

3. Машинное обучение с использованием Python : учебно-методическое пособие / составители А. В. Осин,К. А. Хализев. — Москва :МТУСИ, 2025. — 20 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/501209 (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
- Мультимедийный проектор №1 Casio XJ-V2
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь
- Сетевой монитор:Нулевой клиент Samsung SyncMaster NC240
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Microsoft OfficeProffessionalPlus 2019 Russian
- Microsoft Windows Professional 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)

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Направление и направленность (профиль)
38.04.05 Бизнес-информатика. Информационная бизнес-аналитика

Год набора на ОПОП
2026

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
38.04.05 «Бизнес-информатика» (М-БИ)	ОПК-3 : Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.1к : Использует системы искусственного интеллекта в рамках стратегического планирования и прогнозирования в профессиональной деятельности
		ОПК-3.2к : Использует методы, модели, алгоритмы, технологии и инструментальные средства для работы с большими данными для поддержки стратегических управленческих решений
	ПКВ-2 : Способен выбирать и осуществлять рациональные решения по ИС и ИКТ для управления организациями на основе больших данных	ПКВ-2.1к : Использует и совершенствует методы, модели, алгоритмы, технологии и инструментальные средства для работы с большим и данными
		ПКВ-2.2к : Разрабатывает технические проекты в сфере ИТ, позволяющие анализировать и использовать большие данные в управлении организацией

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 ПКВ-2 «Способен выбирать и осуществлять рациональные решения по ИС и ИКТ для управления организациями на основе больших данных»

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	
ПКВ-2.1к : Использует и совершенствует методы, модели, алгоритмы, технологии и инструментальные средства для работы с большими данными	РД 4	Знание	методов исследования современных проблем и методов прикладной информатики	способен выбрать метод исследования адекватный рассматриваемой проблеме
	РД 5	Умение	разрабатывать оригинальные алгоритмы с использованием языка программирования Python для решения профессиональных задач	способен разработать работоспособный оригинальнйе алгоритм с использованием языка программирования Python
	РД 6	Навык	разработки приложений на языке Python, реализующие методы МО	способен разработать приложение на языке Python, реализующее методы МО, протестировать на реальных данных

ПКВ-2.2к : Разрабатывает технические проекты в сфере ИТ, позволяющие анализировать и использовать большие данные в управлении организацией	РД 1	Знание	аналитических методов обработки статистических данных	выбирает аналитические методы обработки статистических данных соответствующие прикладной задаче
	РД 2	Умение	применять методы машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	применяет методы машинного обучения при решении нестандартных задач реальной практики
	РД 3	Навык	анализа и оценки результатов МО	формирует аналитические отчеты, создает практические рекомендации
	РД 4	Знание	методов исследования современных проблем и методов прикладной информатики	способен выбрать метод исследования адекватный рассматриваемой проблеме

Компетенция ОПК-3 «Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта»

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-3.1к : Использует системы искусственного интеллекта в рамках стратегического планирования и прогнозирования в профессиональной деятельности	РД 1	Знание	аналитических методов обработки статистических данных	выбирает аналитические методы обработки статистических данных соответствующие прикладной задаче
	РД 2	Умение	применять методы машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	применяет методы машинного обучения при решении нестандартных задач реальной практики
	РД 3	Навык	анализа и оценки результатов МО	формирует аналитические отчеты, создает практические рекомендации
ОПК-3.2к : Использует методы, модели, алгоритмы, технологии и инструментальные средства для работы с большими данными для поддержки стратегических управленческих решений	РД 4	Знание	методов исследования современных проблем и методов прикладной информатики	способен выбрать метод исследования адекватный рассматриваемой проблеме
	РД 5	Умение	разрабатывать оригинальные алгоритмы с использованием языка программирования Python для решения профессиональных задач	способен разработать работоспособный оригинальнйе алгоритм с использованием языка программирования Python
	РД 6	Навык	разработки приложений на языке Python, реализующие методы МО	способен разработать приложение на языке Python, реализующее методы МО, протестировать на реальных данных

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 : аналитических методов обработки статистических данных	1.1. Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python	Опрос	Кейс-задача
		1.2. Множественная регрессия и корреляция	Опрос	Кейс-задача
		1.3. Снижение размерности. Метод главных компонент.	Опрос	Кейс-задача
		1.4. Задачи классификации. Модели с дискретной зависимой переменной	Опрос	Кейс-задача
РД2	Умение : применять методы машинного обучения при решении нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	1.1. Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python	Опрос	Кейс-задача
		1.2. Множественная регрессия и корреляция	Опрос	Кейс-задача
		1.3. Снижение размерности. Метод главных компонент.	Опрос	Кейс-задача
		1.4. Задачи классификации. Модели с дискретной зависимой переменной	Опрос	Кейс-задача
		1.5. Кластерный анализ	Опрос	Кейс-задача
РД3	Навык : анализа и оценки результатов МО	1.1. Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python	Опрос	Кейс-задача
		1.2. Множественная регрессия и корреляция	Опрос	Кейс-задача
		1.3. Снижение размерности. Метод главных компонент.	Опрос	Кейс-задача
		1.4. Задачи классификации. Модели с дискретной зависимой переменной	Опрос	Кейс-задача

		1.5. Кластерный анализ	Опрос	Кейс-задача
РД4	Знание : методов исследования современных проблем и методов прикладной информатики	1.3. Снижение размерности. Метод главных компонент.	Опрос	Кейс-задача
		1.5. Кластерный анализ	Опрос	Кейс-задача
РД5	Умение : разрабатывать оригинальные алгоритмы с использованием языка программирования Python для решения профессиональных задач	1.1. Методы машинного обучения. Регрессионный анализ. Парная регрессия. Построение регрессий в RStudio и / или python	Опрос	Кейс-задача
		1.2. Множественная регрессия и корреляция	Опрос	Кейс-задача
		1.3. Снижение размерности. Метод главных компонент.	Опрос	Кейс-задача
		1.4. Задачи классификации. Модели с дискретной зависимой переменной	Опрос	Кейс-задача
		1.5. Кластерный анализ	Опрос	Кейс-задача
РД6	Навык : разработки приложений на языке Python, реализующие методы МО	1.2. Множественная регрессия и корреляция	Опрос	Кейс-задача
		1.4. Задачи классификации. Модели с дискретной зависимой переменной	Опрос	Кейс-задача
		1.5. Кластерный анализ	Опрос	Кейс-задача

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.

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.

Вид учебной деятельности	Оценочное средство		
	опрос	кейс-задача	Итого
Практические занятия	20		20
Промежуточная аттестация	50	30	80
Итого	70	30	100

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.

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 Примеры заданий для выполнения практических работ

После выполнения предусмотренных практических заданий, дял оценивания степени освоения компетенций, проводится Опрос.

Пример вопросов для проведения опросов по результатам занятий

Классические регрессии

1.1. В чём разница между Ridge и Lasso регрессиями?

- Ridge (L2) уменьшает коэффициенты, но не обнуляет (хорош при мультиколлинеарности).

- Lasso (L1) обнуляет некоторые коэффициенты (отбор признаков).

1.2. Почему в логистической регрессии используют сигмоиду?

- Сигмоида преобразует линейную комбинацию в вероятность (0; 1), что интерпретируемо для классификации.

1.3. Как оценить важность признаков в линейной регрессии?

- Через стандартизованные коэффициенты или p-значения (для значимости).

1.4. Что гетероскедастичность и как её проверить?

- Непостоянство дисперсии ошибок. Проверка: тест Уайта, график остатков vs предсказаний.

2. Методы кластеризации

2.1. Почему k-means может давать плохие результаты на данных разной плотности?

- K-means предполагает сферические кластеры одинакового размера, не работает для сложных форм.

2.2. Как выбрать оптимальное число кластеров в k-means?

- Метод локтя (Elbow), силуэт (Silhouette), индекс Дэвиса-Болдуина.

2.3. В чём преимущество DBSCAN перед k-means?

- Находит кластеры произвольной формы и шумовые точки, не требует задания числа кластеров.

2.4. Как работает EM-алгоритм в GMM?

- Итеративно оценивает параметры гауссиан (μ, σ) через максимизацию правдоподобия (E-шаг — оценка скрытых переменных, M-шаг — обновление параметров).

Краткие методические указания

При выполнении задания допускается объединение магистрантов в группы по 2-5 человек. Практические работы позволяет определить уровень усвоения материала. Перед выполнением работы необходимо ознакомиться с теоретическим материалом, проработать методы решения задач, рассмотренных в типовых примерах. После выполнения работы предусмотрен письменный аналитический отчет и публичная защита полученных результатов моделирования, сравнение и обсуждение результатов всех участников

Шкала оценки

,

Оценка	Баллы	Описание
5	17–18	выставляется студенту, если практическое задание выполнено, он лично отвечал на воп росы преподавателя, четко представляя свою позицию и аргументируя точку зрения
4	14–16	выставляется студенту, если практическое задание выполнено, но он лично, отвечая на вопросы преподавателя, недостаточно четко представлял свою позицию и аргументиро вал точку зрения
3	9–13	выставляется студенту, если практическое задание выполнено не полностью, ответы не полные и содержат ошибки
2	0–8	выставляется студент не выполнившему полностью задание

5.2 Задания для решения кейс-задачи

Пример кейс-задачи

Для массива данных «Гаджеты» выявить «знания» об изучаемом объекте:

- выбрать МО с учителем/без учителя, который могут быть использованы для предложенного массива данных – выбор обосновать;
- построить модели для выбранных МО, провести анализ полученных результатов.

По существу, вам нужно ответить на вопросы, и дать пояснения по полученному результату (см. таблицу). Комментарии и ответы давать прямо в скрипте.

вопрос	Правило (rule)	МО, модель и полученные оценки
Можно ли утверждать, что це на гаджета определяется толь ко Брендом и операционной с истемой (OS)?		
Различимы страны производи тели, по техническим количес твенным характеристикам?		
Различимы торговые сети, пр едставленные в Приморском крае, по техническим количес твенным и качественным хара ктеристикам гаджетов?		

Краткие методические указания

Для выполнения кейс-задания студент должен создать скрипт (язык R и/или python) и выполнить следующие шаги:

- загрузить данные, провести анализ качества данных;
- проверить гипотезы о распределении;
- сформулировать гипотезы, формализующие поставленные вопросы;
- выбрать модели МО, позволяющие проверить гипотезы, оценить модели;
- сформудировать полученные знания

Шкала оценки

Оценка	Баллы	Описание
5	36-40	выставляется студенту, если он ответил на более, чем 90 % поставленных вопросов
4	26–35	выставляется студенту, если он ответил на 76-90%% поставленных вопросов
3	11–25	выставляется студенту, если он ответил на 61-75%% поставленных вопросов
2	0–10	выставляется студенту, если он ответил не более, чем на 60% поставленных вопросов,

ФОС и ключи – Машинное обучение(уровень магистратуры)

Перечень вопросов для собеседования по результатам практических работ

1. Классические регрессии

1.1. В чём разница между Ridge и Lasso регрессиями?

- Ridge (L2) уменьшает коэффициенты, но не обнуляет (хорош при мультиколлинеарности).
- Lasso (L1) обнуляет некоторые коэффициенты (отбор признаков).

1.2. Почему в логистической регрессии используют сигмоиду?

- Сигмоида преобразует линейную комбинацию в вероятность (0; 1), что интерпретируемо для классификации.

1.3. Как оценить важность признаков в линейной регрессии?

- Через стандартизованные коэффициенты или p-значения (для значимости).

1.4. Что гетероскедастичность и как её проверить?

- Непостоянство дисперсии ошибок. Проверка: тест Уайта, график остатков vs предсказаний.

2. Методы кластеризации

2.1. Почему k-means может давать плохие результаты на данных разной плотности?

- K-means предполагает сферические кластеры одинакового размера, не работает для сложных форм.

2.2. Как выбрать оптимальное число кластеров в k-means?

- Метод локтя (Elbow), силуэт (Silhouette), индекс Дэвиса-Болдуина.

2.3. В чём преимущество DBSCAN перед k-means?

- Находит кластеры произвольной формы и шумовые точки, не требует задания числа кластеров.

2.4. Как работает EM-алгоритм в GMM?

- Итеративно оценивает параметры гауссиан (μ, σ) через максимизацию правдоподобия (E-шаг — оценка скрытых переменных, M-шаг — обновление параметров).

3. Методы классификации

3.1. Почему SVM устойчив к переобучению в высокомерных пространствах?

- Максимизирует зазор (margin), использует регуляризацию (параметр C).

3.2. Как работает критерий Джини в Decision Tree?

- Минимизирует вероятность ошибки:

$\text{Gini} = 1 - \sum(p_i)^2$, где p_i — доля класса в узле.

3.3. В чём разница между Bagging и Boosting?
- Bagging (Random Forest): параллельное обучение, уменьшение дисперсии.
- Boosting (XGBoost): последовательное, уменьшение смещения, веса ошибок.

3.4. Почему ROC-AUC лучше accuracy при дисбалансе классов?
- AUC оценивает качество разделения классов на всех порогах, а accuracy зависит от порога.

4. Ассоциативные правила
4.1. Как интерпретировать Support и Confidence в Apriori?
- Support: доля транзакций с набором $X \cup Y$
- Confidence: (P(Y|X)) — условная вероятность.

4.2. Почему Apriori неэффективен для больших наборов данных?
- Множество кандидатов, требует много памяти и вычислений (альтернатива: FP-Growth).

4.3. Что такое Lift в ассоциативных правилах?
- Мера зависимости:

$$\text{Lift} = \frac{P(X \cap Y)}{P(X)P(Y)}.$$

Lift > 1 — полезное правило.

4.4. Как бороться с ложными ассоциациями?
- Использовать Lift, корреляцию или тесты на значимость (χ^2).

Дополнительные сложные вопросы
- Про ансамбли: Как работает градиентный бустинг? Почему LightGBM быстрее XGBoost?
- Про нейросети: Как избежать переобучения в CNN? Что делает Batch Normalization?
- Про интерпретацию: Как SHAP объясняет предсказания модели?

5. Как SHAP объясняет предсказания модели?
SHAP — это метод объяснения предсказаний модели на основе теории игр (Shapley values). Он количественно оценивает вклад каждого признака в конкретное предсказание, обеспечивая интерпретируемость даже для сложных моделей (например, градиентного бустинга или нейросетей). Значения SHAP показывают, насколько признак увеличил или уменьшил прогноз относительно среднего значения по выборке.
Развёрнутое объяснение (если нужны детали):

1. **Основа метода**:
    - SHAP использует **значения Шепли** из теории кооперативных игр: вклад каждого признака вычисляется как средневзвешенное влияние на прогноз по всем возможным комбинациям признаков.
    - Формула для Shapley value:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} (f(S \cup \{i\}) - f(S))$$

где F — все признаки, S — подмножество признаков, f — модель.

2. **Свойства SHAP**:
   - **Локальная интерпретация**: Объясняет индивидуальные предсказания (например, почему кредит заявке отказано).
   - **Глобальная интерпретация**: Агрегация значений SHAP по выборке показывает общую важность признаков.
   - **Аддитивность**: Сумм значений SHAP равна разнице между предсказанием и средним значением модели ($\sum \phi i = f(x) - E[f(x)]$).
3. **Практическое применение**:
   - Визуализация:
     - **Force plot**: Показывает вклад каждого признака для одного объекта.
     - **Summary plot**: Ранжирует признаки по важности на всей выборке.
   - Пример: В кредитном скоринге SHAP может показать, что высокий долг ($\phi = +0.3$) увеличил вероятность дефолта, а большой стаж работы ($\phi = -0.2$) — уменьшил.
4. **Плюсы и минусы**:
   - **Плюсы**:
     - Работает для любых моделей (black-box).
     - Соответствует человеческой интуиции («вклад в результат»).
   - **Минусы**:
     - Вычислительно сложен для больших datasets (ускоренные версии: KernelSHAP, TreeSHAP).
     - Может быть нестабильным при сильной корреляции признаков.

Что могут спросить дополнительно?

- **Чем SHAP отличается от LIME?**
  - LIME аппроксимирует модель локально линейной, а SHAP даёт точные значения вклада на основе теории игр.
- **Как ускорить SHAP для ансамблей деревьев?**
  - Использовать **TreeSHAP** (оптимизированную версию с полиномиальной сложностью).

Ключи для кейс-задания

1. Для массива данных «Гаджеты (Часы)» выявить «знания» об изучаемом объекте:

- выбрать МО с учителем/без учителя, который могут быть использованы для предложенного массива данных – выбор обосновать;
- построить модели для выбранных МО, провести анализ полученных результатов.

вопрос	Правило (rule)	МО, модель и полученные оценки
Можно ли утверждать, что цена гаджета	Бренд и ОС важные признаки	Линейная регрессия: R^2: 0.62 – Бренд и ОС

определяется только Брендом и операционной системой (OS)?	ценообразования, но на цену также влияют тех. Характеристики, память, модель, размер и материал корпуса и тд.	R^2: 0.93 – Тех. Характ.
Различимы страны производители, по техническим количественным характеристикам?	Страны различаются. При помощи кластеризации выделили 3 группы на основе Тех. Характ.	KMeans: Получено 3 кластера с различием по объёму памяти, разрешению экрана и размеру корпуса.
Различимы торговые сети, представленные в Приморском крае, по техническим количественным и качественным характеристикам гаджетов?	Торговые сети различимы, Основные различия: Разрешение экрана Размер корпуса Обьем ОЗУ (ГБ) Обьем встроенной памяти (ГБ)	Случайный лес: MAE: 925.94 MSE: 1818992.99 R^2: 0.97 Основные признаки: Разрешение экрана 0.387362 Размер корпуса 0.268570 Обьем ОЗУ (ГБ) 0.244277 Обьем встроенной памяти (ГБ) 0.099792