

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»
(СибГУТИ)
Уральский технический институт связи и информатики (филиал) в г. Екатеринбурге
(УрТИСИ СибГУТИ)

УТВЕРЖДАЮ
директор УрТИСИ СибГУТИ
Минина Е.А.
« 23 » 11 2025 г.

ОЦЕНОЧНЫЕ СРЕДСТВА ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПО ДИСЦИПЛИНЕ

Б1.В.ДВ.01.01 Машинное обучение и аналитика данных

Направление подготовки / специальность: **09.03.01 «Информатика и
вычислительная техника»**

Направленность (профиль) / специализация: **Инженерия программного
обеспечения и искусственного интеллекта**

Форма обучения: **очная**

Год набора: 2026

Разработчик (-и):
преподаватель

подпись

/А.Е. Каменсков/

к.т.н. доцент

подпись

/ Н.В.Будылдина /

Оценочные средства обсуждены и утверждены на заседании кафедры инфокоммуникационных технологий и мобильной связи (ИТиМС)

Протокол от 27.11.2025 г. № 3

Заведующий кафедрой _____ / Н.В.Будылдина /
подпись

Екатеринбург, 2025

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»
(СибГУТИ)
Уральский технический институт связи и информатики (филиал) в г. Екатеринбурге
(УрТИСИ СибГУТИ)

УТВЕРЖДАЮ
директор УрТИСИ СибГУТИ
_____ Минина Е.А.
« ____ » _____ 2025 г.

ОЦЕНОЧНЫЕ СРЕДСТВА ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПО ДИСЦИПЛИНЕ

Б1.В.ДВ.01.01 Машинное обучение и аналитика данных

Направление подготовки / специальность: **09.03.01 «Информатика и
вычислительная техника»**

Направленность (профиль) / специализация: **Инженерия программного
обеспечения и искусственного интеллекта**

Форма обучения: **очная**

Год набора: 2026

Разработчик (-и):
преподаватель

_____ /А.Е. Каменсков/
подпись

К.Т.Н. доцент

_____ / Н.В.Будылдина /
подпись

Оценочные средства обсуждены и утверждены на заседании кафедры инфокоммуникационных технологий и мобильной связи (ИТиМС)

Протокол от 27.11.2025 г. № 3

Заведующий кафедрой _____ / Н.В.Будылдина /
подпись

Екатеринбург, 2025

1. Перечень компетенций и индикаторов их достижения

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенций	Этап	Предшествующие этапы (с указанием дисциплин/практик)
ПК-5 Способен применять математические методы для моделирования и разработки систем сбора, анализа и обработки данных с использованием современных информационных технологий	ПК-5.1 Знает математические методы и средства сбора, анализа и обработки данных	1	-
	ПК-5.2 Умеет применять математические методы и средства сбора, анализа и обработки данных		
	ПК-5.3 Владеет навыками применения математических методов для моделирования и разработки систем сбора, анализа и обработки данных с использованием современных информационных технологий		

Форма промежуточной аттестации по дисциплине – экзамен

2. Показатели, критерии и шкалы оценивания компетенций

2.1 Показателем оценивания компетенций на этапе их формирования при изучении дисциплины является уровень их освоения.

Индикатор освоения компетенции	Показатель оценивания	Критерий оценивания
ПК-5.1 Знает математические методы и средства сбора, анализа и обработки данных	Знает математические методы и средства сбора, анализа, обработки данных в машинном обучении и обработки данных.	Демонстрирует уверенные знания о принципах машинного обучения и аналитике данных на экзамене.
ПК-5.2 Умеет применять математические методы и средства сбора, анализа и обработки данных	Умеет математические методы и средства сбора, анализа, обработки данных в машинном обучении и обработки данных.	Демонстрирует уверенные умения составления программ машинного обучения и аналитике данных на экзамене.
ПК-5.3 Владеет навыками применения	Владеет навыками применения	Демонстрирует уверенное владение полного цикла обработки данных и

математических методов для моделирования и разработки систем сбора, анализа и обработки данных с использованием современных информационных технологий	математических методов для моделирования и разработки систем сбора, анализа и обработки данных с использованием современных информационных технологий в машинном обучении и обработки данных.	составлений программ машинного обучения.
---	---	--

Шкала оценивания.

Экзамен

5-балльная шкала	Критерии оценки
«отлично»	На экзаменационные вопросы даны полные аргументированные ответы. На экзаменационные вопросы даны полные, аргументированные ответы с приведением примеров. Студент демонстрирует сформированность дисциплинарных компетенций на итоговом уровне, обнаруживает всестороннее, систематическое и глубокое знание учебного материала по всей тематике курса: фундаментальные основы, предобработка данных и разведочный анализ (EDA), Обучение с учителем, Обучение без учителя, Нейронные сети, Аналитика данных. Студент усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, свободно оперирует приобретенными знаниями, умениями, применяет их при выполнении заданий.
«хорошо»	На экзаменационные вопросы даны полные аргументированные ответы, но с замечаниями преподавателя. Студент демонстрирует сформированность дисциплинарных компетенций на среднем уровне: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при ответе на поставленные вопросы, по тематике фундаментальные основы, предобработка данных и разведочный анализ (EDA), Обучение с учителем, Обучение без учителя, Нейронные сети, Аналитика данных. Допущены ошибки при решении задач
«удовлетворительно»	На экзаменационные вопросы даны ответы со слабой аргументацией, преподаватель задал множество наводящих вопросов. Студент демонстрирует сформированность дисциплинарных компетенций на базовом уровне: в ходе выполнения практических заданий, решения задач допускаются значительные ошибки, проявляется отсутствие отдельных знаний, по некоторым дисциплинарным разделам, студент испытывает значительные затруднения при оперировании знаниями и по тематике: фундаментальные основы, предобработка данных и разведочный анализ (EDA), Обучение с учителем, Обучение без учителя, Нейронные сети, Аналитика данных.

«неудовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций на уровне ниже порогового, проявляется недостаточность знаний. Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний по темам дисциплины, отсутствуют навыки решения задач.
-----------------------	---

3. Методические материалы, определяющие процедуры оценивания по дисциплине

3.1. В ходе реализации дисциплины используются следующие формы и методы текущего контроля

Тема и/или раздел	Формы/методы текущего контроля успеваемости
ПК-5.1 Знает математические методы и средства сбора, анализа и обработки данных	
Введение и фундаментальные основы	Лекции
Знакомство с инструментами и Python для анализа.	Практические занятия
Визуализация и статистика на Python.	Практические занятия
Глубокий разведочный анализ (EDA) реального набора данных	Практические занятия
ПК-5.2 Умеет применять математические методы и средства сбора, анализа и обработки данных	
Feature Engineering и пайплайн предобработки.	Практические занятия
Линейные модели для регрессии и классификации.	Практические занятия
Ансамбли моделей и метрики.	Практические занятия
SVM, k-NN и валидация моделей.	Практические занятия
Кластеризация и интерпретация результатов.	Практические занятия
Снижение размерности и поиск аномалий	Практические занятия
ПК-5.3 Владеет навыками применения математических методов для моделирования и разработки систем сбора, анализа и обработки данных с использованием современных информационных технологий	
Построение нейросетей на PyTorch/TensorFlow	Практические занятия
Сверточные нейронные сети (CNN)	Практические занятия
Рекуррентные нейронные сети (RNN) для работы с последовательностями.	Практические занятия
Комплексный анализ и разработка предсказательной модели	Практические занятия

3.2. Типовые материалы текущего контроля успеваемости обучающихся

ОПК-5 Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем

Пример задания на практическое занятие

1 Цель работы:

1.1 Изучить основы работы с методами классификации в машинном обучении.

2 Подготовка к работе:

2.1 Используя литературу, изучить работу с методами машинного обучения и библиотекой Tensorflow.

3 Задание:

3.1 Необходимо перейти по ссылке, просмотреть пример и выполнить следующее задание, используя метод kNN.

Исходные данные:

Всего 5 классов, каждый класс должен иметь свой уникальный цвет, на каждый класс по 30 объектов (можете использовать свои данные, главное задокументировать)

Ссылка на пример метода kNN:
https://colab.research.google.com/drive/1ULIFjTDe1DNCHPLc0bu_bbatU-HNbHJJ?usp=sharing

3.2 Необходимо перейти по ссылке, просмотреть пример и выполнить следующее задание, используя метод решающих деревьев.

Исходные данные:

Представьте себе, что вы - медицинский исследователь, собирающий данные для исследования. Вы собрали данные о нескольких пациентах, страдающих одним и тем же заболеванием. В ходе лечения каждый пациент принимал один из 5 препаратов: препарат А, препарат В, препарат С, препараты X и Y.

Ваша задача состоит в том, чтобы построить модель, позволяющую определить, какое лекарство может подойти будущему пациенту с тем же заболеванием. Признаками этого набора данных являются возраст, пол, артериальное давление и уровень холестерина у пациентов, а целью - лекарство, на которое отреагировал каждый пациент.

Это пример многоклассового классификатора, и вы можете использовать обучающую часть набора данных для построения дерева решений, а затем использовать его для предсказания класса неизвестного пациента или для назначения лекарства новому пациенту.

Источник данных: IBM

Ссылка на набор данных:
<https://drive.google.com/file/d/11GtR4R4aRHvASNhkT3nvtNrUaJtCCi4D/view?usp=sharing>

Ссылка на пример метода решающих деревьев и на набор данных:
<https://colab.research.google.com/drive/1wSy3C-SuqC9yJFex3hPAvXNWDnhYUSdc#scrollTo=Rm9E8HBmUtuH>
<https://drive.google.com/file/d/1gLp4K7SFuWsM5rcD3MTWTIPSnwkomOW/view?usp=sharing>

3.3 Необходимо перейти по ссылке, просмотреть пример и выполнить следующее задание, используя логическую регрессию.

Исходные данные:

Сделать оценку набора данных о выживаемости "Титаника" при помощи логической регрессии.

Ссылка на набор данных:

<https://drive.google.com/drive/folders/10E3QfSH7heG1Z62eA5KpmNvLMrKkKaOb?usp=sharing>

Ссылка на пример метода логической регрессии:
<https://colab.research.google.com/drive/1WpbVl5eTdCtcq0og86xWVuCst72ZB9mF?usp=sharing>

3.3. Типовые материалы для проведения промежуточной аттестации обучающихся

Типовые вопросы и задания к экзамену:

1. Дайте определение машинного обучения. В чем ключевое отличие обучения с учителем, без учителя и с подкреплением? Приведите по два примера реальных задач для каждого типа.

2. Опишите основные этапы жизненного цикла проекта машинного обучения (например, по методологии CRISP-DM). Какой этап, на ваш взгляд, наиболее критичен и почему?

3. Что такое переобучение (overfitting) и недообучение (underfitting)? Опишите визуальное поведение ошибок на тренировочной и валидационной выборках в процессе обучения модели. Какие существуют методы борьбы с переобучением?

4. В чем разница между ковариацией и коэффициентом корреляции (Пирсона)? Как корреляция связана с линейной регрессией? Почему наличие корреляции не означает причинно-следственную связь?

5. Что такое разведочный анализ данных (EDA) и каковы его основные цели? Опишите три ключевых действия или типа визуализации, которые вы выполняете на этапе EDA, и какую информацию они дают.

6. Объясните разницу между масштабированием (StandardScaler) и нормализацией (MinMaxScaler). В каких случаях предпочтительнее каждое из них? Всегда ли это необходимо для деревьев решений?

7. Какие существуют основные стратегии работы с пропущенными значениями (missing values)? В чем разница между удалением строк (dropna) и импутацией (SimpleImputer)? Когда уместно использовать каждую?

8. Что такое Feature Engineering? Приведите три примера создания новых признаков из существующих (для числовых и категориальных данных) и объясните, как это может улучшить модель.

9. Опишите математическую постановку задачи линейной регрессии. Что минимизирует метод наименьших квадратов (MSE)? В чем смысл коэффициентов (весов) модели?

10. Объясните концепцию регуляризации. В чем разница между L1- (Lasso) и L2-регуляризацией (Ridge) с точки зрения математики, влияния на веса модели и интерпретируемости?

11. Как работает логистическая регрессия для задачи бинарной классификации? Что представляет собой ее выход и какова используемая функция потерь (Log Loss)?

12. Объясните принцип работы метода k-ближайших соседей (k-NN). Как выбор метрики расстояния и параметра k влияет на результат? В чем главный недостаток этого алгоритма?

13. В чем заключается идея метода опорных векторов (SVM) для линейно разделимых данных? Что такое разделяющая гиперплоскость, опорные векторы и зазор (margin)? Зачем нужен «трюк с ядром» (kernel trick)?

14. Объясните разницу между бэггингом (Bagging) и бустингом (Boosting). Как работает алгоритм Random Forest (как частный случай бэггинга) и почему он менее склонен к переобучению, чем одно дерево?

15. Опишите принцип работы градиентного бустинга (Gradient Boosting) «в одну строку». Назовите три популярные современные реализации и их ключевые особенности (например, XGBoost, LightGBM).

16. Для задачи классификации назовите не менее 5 метрик качества, объясните их формулы и интерпретацию. В каких практических ситуациях Precision важнее Recall, и наоборот? Что показывает ROC-AUC кривая?

17. Что такое кросс-валидация (k-Fold CV) и зачем она нужна? В чем разница между GridSearchCV и RandomizedSearchCV для подбора гиперпараметров?

18. Почему случайный лес часто называют «моделью по умолчанию» для табличных данных? Назовите его сильные и слабые стороны по сравнению с градиентным бустингом.

19. Опишите алгоритм K-Means. Как выбирается оптимальное число кластеров (метод локтя, силуэтный коэффициент)? Каковы основные предпосылки и ограничения этого алгоритма?

20. В чем принципиальное отличие алгоритма кластеризации DBSCAN от K-Means? Какие параметры ему нужны и что такое точки «ядра», «границы» и «шума»?

21. Для чего используется метод главных компонент (PCA)? Объясните, что такое «объясненная дисперсия». Является ли PCA методом feature selection (отбора признаков)?

22. Из каких основных компонентов состоит искусственный нейрон? Назовите и изобразите графики трех наиболее распространенных функций активации (ReLU, sigmoid, tanh) и объясните их свойства.

23. Опишите алгоритм обратного распространения ошибки (Backpropagation). Какую роль в нем играет цепное правило (chain rule) дифференцирования и градиентный спуск?

24. В чем ключевые архитектурные особенности сверточных нейронных сетей (CNN), делающие их эффективными для обработки изображений? Объясните роль операций свертки (convolution) и пулинга (pooling).

25. Почему простые RNN плохо справляются с долгосрочными зависимостями? Как архитектуры LSTM и GRU решают эту проблему с помощью механизмов «вентилей» (gates)?

26. Назовите три современных оптимизатора для обучения нейронных сетей, кроме стохастического градиентного спуска (SGD) (например, Adam). В чем их основное преимущество?

27. В чем ключевые философские и практические различия между фреймворками PyTorch и TensorFlow (Keras)? Какой из них вы бы выбрали для быстрого прототипирования и почему?

28. Представьте, что вы обучили модель, которая показывает отличные результаты на тестовой выборке, но проваливается при развертывании в реальной системе. Назовите три возможные причины такой ситуации (помимо ошибки в коде).

29. Что такое A/B-тестирование в контексте внедрения ML-модели? Какие метрики и статистические критерии используются для сравнения старой и новой системы?

30. Опишите базовые шаги для создания простого продакшн-пайплайна машинного обучения: от сохранения обученной модели (pickle, joblib, .pth) до создания минимального API-эндпоинта (например, на Flask/FastAPI). Что такое сериализация (serialization) модели и зачем она нужна?

Банк контрольных вопросов, заданий и иных материалов, используемых в процессе процедур текущего контроля и промежуточной аттестации находится в учебно-методическом комплексе дисциплины и/или представлен в электронной информационно-образовательной среде по URI: <http://www.aup.uisi.ru>.

3.4. Методические материалы проведения текущего контроля и промежуточной аттестации обучающихся

Перечень методических материалов для подготовки к текущему контролю и промежуточной аттестации:

1. Методические указания по выполнению практических занятий по дисциплине «Машинное обучение и аналитика данных». –URL: <https://aup.uisi.ru/4040330/>

2. Методические указания по выполнению лабораторных работ по дисциплине «Машинное обучение и аналитика данных». –URL: <https://aup.uisi.ru/4040330/>