МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

УТВЕРЖДАЮ
Проректор по учебной работе
_____ М.А. Смирнов
« _____ » _____ 2026 г.

РАБОЧАЯ ПРОГРАММА

дисциплины части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)»

«Машинное обучение и большие данные»

Направление подготовки бакалавров – 09.03.01. Информатика и вычислительная техника

Направленность (профиль) - Вычислительные машины, комплексы, системы и сети

Типы задач профессиональной деятельности: производственно-технологический

Форма обучения – очная, заочная ускоренная

Факультет информационных технологий
Кафедра электронных вычислительных машин

Тверь 2026

Рабочая программа дисциплины соответствует ОХОП подготовки бакалавров в части требований к результатам обучения по дисциплине и учебному плану.

Разработчик программы: к.т.н., доцент Е.Е. Фомина

Программа рассмотрена и одобрена на заседании кафедры ЭВМ
«31» 10 2025 г., протокол № 2.

Заведующий кафедрой А.Р. Хабаров

Согласовано
Начальник учебно-методического
отдела УМУ Е.Э. Наумова

Начальник отдела
комплектования
зональной научной библиотеки О.Ф. Жмыхова

1. Цель и задачи дисциплины

Целью изучения дисциплины «Машинное обучение и большие данные» является формирование у обучающихся компетенций, необходимых для всесторонней работы с большими объемами структурированной и неструктурированной информации, включая сбор, разработку и применение моделей данных для извлечения новых знаний и обеспечения эффективной информационно-аналитической поддержки принятия обоснованных управленческих решений.

Задачами дисциплины являются:

формирование у обучающихся системного понимания фундаментальных принципов машинного обучения и архитектуры систем обработки больших данных;

освоение методики эффективной подготовки, очистки, трансформации и валидации больших массивов структурированных и неструктурированных данных;

приобретение практических навыков по разработке, настройке и верификации прогностических и классификационных модели данных, используя современные алгоритмы машинного обучения;

развитие навыков интерпретации результатов, полученных с помощью моделей машинного обучения, и представления этих результатов в форме, пригодной для информационно-аналитической поддержки принятия управленческих решений;

приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;

приобретение практических навыков работы большими данными.

2. Место дисциплины в структуре ООП

Дисциплина относится к части, формируемой участниками образовательных отношений, Блока 1 ОП ВО. Для изучения курса требуются знания, умения и навыки, полученные в процессе изучения дисциплин «Информатика», «Алгоритмические языки и программирование», «Теория вероятностей, математическая статистика и случайные процессы», «Вычислительная математика» и «Математика».

Приобретенные знания в рамках данной дисциплины помимо их самостоятельного значения являются основой для изучения курсов «Системы искусственного интеллекта», «Теория принятия решений» и других дисциплин, профессиональная подготовка по которым предполагает использование технологий анализа больших данных при решении задач, а также при выполнении выпускной квалификационной работы.

3. Планируемые результаты обучения по дисциплине

3.1. Планируемые результаты обучения по дисциплине

Компетенция, закрепленная за дисциплиной в ОХОП:

УК-2. *Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений*.

Индикатор компетенции, закреплённый за дисциплиной в ОХОП:

ИУК-2.2. Выбирает оптимальный способ решения задач, учитывая действующие правовые нормы и имеющиеся условия, ресурсы и ограничения.

Показатели оценивания индикатора достижения компетенции:

ИУК-2.2:

Знать:

З1.1: Алгоритм выбора оптимального способа решения задач.

Уметь:

У1.1: Анализировать имеющиеся ресурсы для решения поставленной задачи.

Иметь опыт практической подготовки:

ПП1.1: Опыт по постановке цели проведения работ, выделению необходимых задач, для решения поставленной цели, и по грамотному обоснованию принимаемых решений.

ПК-2. *Способен производить разработку компонентов системных программных продуктов, интегрировать программные модули и компоненты, проводить верификацию программного продукта*.

Индикаторы компетенции, закреплённые за дисциплиной в ОХОП:

ИПК-2.1. Разрабатывает системные утилиты.

ИПК-2.2. Анализирует требования к программному обеспечению.

ИПК-2.3. Применяет на практике методы объектно-ориентированного проектирования при разработке программных систем.

Показатели оценивания индикаторов достижения компетенции:

ИПК-2.1:

Знать:

З2.1. Синтаксис и основные функции ключевых библиотек для анализа данных.

З2.2. Стандартные методики предобработки данных (работа с пропущенными значениями, кодирование категориальных признаков).

З2.3. Основные компоненты фреймворков глубокого обучения.

Уметь:

У2.1. Применять программные инструменты для быстрой очистки, трансформации и визуализации больших наборов данных.

У2.2. Использовать стандартные API библиотек для обучения и оценки моделей классификации и кластеризации.

У2.3. Конфигурировать и обучать базовые архитектуры нейронных сетей с использованием специализированных программных средств.

Иметь опыт практической подготовки:

ПП2.1: По разработке аналитического программного обеспечения для анализа данных.

ИПК-2.2:

Знать:

З3.1. Логику и последовательность шагов в алгоритмах кластеризации (K-means, иерархические методы) и классификации (деревья решений, SVM, нейронные сети).

З3.2. Принципы построения ансамблей (Bagging, Boosting).

Уметь:

У3.1. Определять тип задачи (регрессия, классификация, кластеризация) на основе постановки проблемы.

У3.2. Разрабатывать структурированный алгоритм обработки данных, включающий предобработку, выбор модели, обучение и валидацию.

У3.3. Обосновывать выбор конкретного алгоритма машинного обучения на основе типа, объема и структуры исходных данных.

Иметь опыт практической подготовки:

ППЗ.1: По проектированию систем программного обеспечения согласно поставленным требованиям и по выявлению достоинств и недостатков в предоставляемых системах программного обеспечения для их дальнейшей модернизации.

ИПК-2.3:

Знать:

З4.1. Основные принципы статистического анализа данных.

З4.2. Математические предпосылки и ограничения ключевых алгоритмов машинного обучения.

З4.3. Методологию проведения контролируемых экспериментов (разделение выборок, кросс-валидация).

Уметь:

У4.1. Выбирать и применять адекватные метрики для количественной оценки результатов моделирования и кластеризации.

У4.2. Проводить сравнительный анализ различных моделей машинного обучения на одних и тех же наборах данных.

У4.3. Интерпретировать полученные результаты с точки зрения достоверности и статистической значимости.

Иметь опыт практической подготовки:

ПП4.1: По разработке систем программного обеспечения, по составлению проектной и технической документации для программных систем, по реализации готовых решений согласно предоставляемой документации.

3.2. Технологии, обеспечивающие формирование компетенций

Проведение лекционных, лабораторных занятий, самостоятельная работа.

4. Трудоемкость дисциплины и виды учебной работы

ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 1а. Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачетные единицы	Академические часы
Общая трудоемкость дисциплины	5	**180**
Аудиторные занятия (всего)		**90**
В том числе:		
Лекции		45
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		45
Самостоятельная работа обучающихся (всего)		**54+36 (экз)**
В том числе:		
Курсовая работа		20
Курсовой проект		не предусмотрен
Расчетно-графические работы		не предусмотрены
Другие виды самостоятельной работы: - подготовка к лабораторным работам		20
Текущий контроль успеваемости и промежуточная аттестация (зачет)		7
Текущий контроль успеваемости и промежуточная аттестация (экзамен)		7+36(экз)
Практическая подготовка при реализации дисциплины (всего)		65
В том числе:		
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		45
Курсовая работа		20
Курсовой проект		не предусмотрен

ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 1б. Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачетные единицы	Академические часы
Общая трудоемкость дисциплины	4	180
Аудиторные занятия (всего)		14
В том числе:		
Лекции		8
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		6
Самостоятельная работа обучающихся (всего)		153+4 (зач)+9 (экз)
В том числе:		
Курсовая работа		50
Курсовой проект		не предусмотрен
Расчетно-графические работы		не предусмотрены

Вид учебной работы	Зачетные единицы	Академические часы
Другие виды самостоятельной работы: - подготовка к лабораторным работам		50
Текущий контроль успеваемости и промежуточная аттестация (зачет)		25+4 (зач)
Текущий контроль успеваемости и промежуточная аттестация (зачет)		28+9 (экз)
Практическая подготовка при реализации дисциплины (всего)		56
В том числе**:**		
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		6
Курсовая работа		50
Курсовой проект		не предусмотрен

5. Структура и содержание дисциплины.

5.1. Структура дисциплины

ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2а. Модули дисциплины, трудоемкость в часах и виды учебной работы

№	Наименование модуля	Труд-ть часы	Лекции	Практич. занятия	Лаб. работы	Сам. работа
1	Основные понятия машинного обучения	25	4	-	6	*9+6(экз)*
2	Методы машинной кластеризации	27	6	-	6	*9+6(экз)*
3	Методы машинной классификации	37	10	-	12	*9+6(экз)*
4	Нейросетевые технологии	27	6	-	6	*9+6(экз)*
5	Приложения методов машинного обучения	27	6	-	6	*9+6(экз)*
6	Технологии работы с большими данными	37	13	-	9	*9+6(экз)*
Всего на дисциплину		***180***	*45*	-	*45*	*54+36(экз)*

ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2б. Модули дисциплины, трудоемкость в часах и виды учебной работы

№	Наименование модуля	Труд-ть часы	Лекции	Практич. занятия	Лаб. работы	Сам. работа
1	Основные понятия машинного обучения	29	1	-	1	*25+2(экз)*
2	Методы машинной кластеризации	31	1	-	1	*25+4(экз)*
3	Методы машинной классификации	34	2	-	1	*28+3(экз)*
4	Нейросетевые технологии	28	1	-	1	*25+1(экз)*
5	Приложения методов машинного обучения	28	1	-	1	*25+1(экз)*
6	Технологии работы с большими данными	30	2	-	1	*25+2(экз)*
Всего на дисциплину		***180***	*8*	-	*6*	153+4 (зач)+9 (экз)

5.2. Содержание дисциплины

МОДУЛЬ 1 «Основные понятия машинного обучения»

Содержание темы: Системы искусственного интеллекта. Связь машинного обучения с системами искусственного интеллекта. Основные модели машинного обучения. Создание обучающей выборки. Типы выборок. Переобучение.

МОДУЛЬ 2 «Методы машинной кластеризации»

Кластерный анализ. Сущность кластерного анализа и его геометрическая интерпретация. Основные этапы кластерного анализа: выбор системы признаков, снижение размерности признакового пространства, выбор меры близости, выбор алгоритма классификации и оценка её качества. Классификация методов кластерного анализа. Эвристические алгоритмы классификации. Иерархические алгоритмы классификации. Метод к-средних. Критерии качества классификации.

МОДУЛЬ 3 «Методы машинной классификации»

Понятие классификации. Ошибки классификатора. Логические методы классификации. Байесовский классификатор. Деревья классификации. Алгоритмы построения деревьев классификации: CART, CHAID. Random Forest. Логистическая регрессия. Дискриминантный анализ. Решение задач классификации. Оценка качества классификатора. A/B тестирование

МОДУЛЬ 4 «Нейросетевые технологии»

Однослойные нейронные сети. Модель нейрона. Архитектуры нейронных сетей. Основы глубокого обучения. Многослойные нейронные сети. Свёртка. Свёрточные нейронные сети. Рекуррентные нейронные сети.

МОДУЛЬ 5 «Приложения методов машинного обучения»

Тематическое моделирование с использованием методов машинного обучения: выявление трендов в данных, поиск информации, анализ информационных потоков, анализ биоинформационных данных, рубрикация документов. Метод Reinforcement Learning. Приложения использования методов машинного обучения: методы машинного обучения в автомобильной промышленности, автомобили с системами автопилота, робототехника, персональные умные ассистенты.

МОДУЛЬ 6 «Технологии работы с большими данными»

Введение в обработку больших данных. Основные принципы Hadoop. Архитектура Hadoop. Компоненты Hadoop: HDFS, MapReduce, YARN. Установка и конфигурация Hadoop. Работа с Hadoop Distributed File System (HDFS). Запуск MapReduce задач. Примеры использования Hadoop для обработки больших данных.

5.3. Лабораторные работы

ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 3а. Лабораторные работы и их трудоемкость

Порядковый номер модуля. Цели лабораторных работ	**Наименование лабораторных работ**	**Трудоемкость в часах**
Модуль 1. Основные понятия машинного обучения **Цель:** Освоить основные этапы конвейера машинного обучения (сбор, предобработка, обучение, оценка)	Знакомство с инструментарием (Pandas, NumPy, Scikit-learn) и первичный анализ данных. Предварительная обработка данных и стандартизация признаков. Исследование влияния гиперпараметров на качество модели (на примере регуляризации).	**6**
Модуль 2. Методы машинной кластеризации **Цель:** Освоить методы кластеризации, научиться применять различные алгоритмы кластеризации, анализировать результаты и выбирать оптимальный алгоритм для конкретных задач, оценивать качество кластеризации.	Иерархическая кластеризация Кластеризация методом к-средних	**6**
Модуль 3 Методы машинной классификации **Цель:** Изучить методы классификации, научиться применять различные алгоритмы классификации для решения практических задач, оценивать качество классификатора	Байесовский классификатор. Деревья классификации. Random Forest. Логистическая регрессия. Дискриминантный анализ. A/B тестирование	**12**
Модуль 4 Нейросетевые технологии **Цель:** Изучить основы нейронных сетей, освоить архитектуры нейронных сетей, такие как многослойные перцептроны, сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN), научиться решать задачи классификации и регрессии	Разработка НС для решения задачи классификации Разработка НС для решения задачи регрессии	**6**
Модуль 5 Приложения методов машинного обучения **Цель:** Применить методы машинного обучения для решения практических задач, таких как распознавание изображений, обработка естественного языка (NLP), прогнозирование временных рядов, анализ данных.	Применение методов машинного обучения для решения практических задач	**6**

Порядковый номер модуля. Цели лабораторных работ	Наименование лабораторных работ	Трудоемкость в часах
Модуль 6 Технологии работы с большими данными **Цель:** Изучить инструменты и технологии работы с большими данными, такие как Hadoop, Spark, научиться обрабатывать и анализировать большие объемы данных, работать с распределенными вычислениями.	Технологии работы с большими данными	**9**

ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 3б. Лабораторные работы и их трудоемкость

Порядковый номер модуля. Цели лабораторных работ	Наименование лабораторных работ	Трудоемкость в часах
Модуль 1. Основные понятия машинного обучения **Цель:** Освоить основные этапы конвейера машинного обучения (сбор, предобработка, обучение, оценка)	Знакомство с инструментарием (Pandas, NumPy, Scikit-learn) и первичный анализ данных. Предварительная обработка данных и стандартизация признаков. Исследование влияния гиперпараметров на качество модели (на примере регуляризации).	**1**
Модуль 2. Методы машинной кластеризации **Цель:** Освоить методы кластеризации, научиться применять различные алгоритмы кластеризации, анализировать результаты и выбирать оптимальный алгоритм для конкретных задач, оценивать качество кластеризации.	Иерархическая кластеризация Кластеризация методом к-средних	**1**
Модуль 3 Методы машинной классификации **Цель:** Изучить методы классификации, научиться применять различные алгоритмы классификации для решения практических задач, оценивать качество классификатора	Байесовский классификатор. Деревья классификации. Random Forest. Логистическая регрессия. Дискриминантный анализ. A/B тестирование	**1**
Модуль 4 Нейросетевые технологии **Цель:** Изучить основы нейронных сетей, освоить архитектуры нейронных сетей, такие как многослойные перцептроны, сверточные нейронные сети (CNN), рекуррентные нейронные сети	Разработка НС для решения задачи классификации Разработка НС для решения задачи регрессии	**1**

Порядковый номер модуля. Цели лабораторных работ	Наименование лабораторных работ	Трудоемкость в часах
(RNN), научиться решать задачи классификации и регрессии		
Модуль 5 Приложения методов машинного обучения **Цель:** Применить методы машинного обучения для решения практических задач, таких как распознавание изображений, обработка естественного языка (NLP), прогнозирование временных рядов, анализ данных.	Применение методов машинного обучения для решения практических задач	**1**
Модуль 6 Технологии работы с большими данными **Цель:** Изучить инструменты и технологии работы с большими данными, такие как Hadoop, Spark, научиться обрабатывать и анализировать большие объемы данных, работать с распределенными вычислениями.	Технологии работы с большими данными	**1**

5.4. Практические и (или) семинарские занятия

Учебным планом практические занятия не предусмотрены.

6. Самостоятельная работа обучающихся и текущий контроль успеваемости

6.1. Цели самостоятельной работы

Формирование способностей к самостоятельному познанию и обучению, поиску литературы, обобщению, оформлению и представлению полученных результатов, их критическому анализу, поиску новых и неординарных решений, аргументированному отстаиванию своих предложений, умений подготовки выступлений и ведения дискуссий.

6.2. Организация самостоятельной работы

Самостоятельная работа заключается в изучении отдельных тем курса по заданию преподавателя по рекомендуемой им учебной литературе, в подготовке к лабораторным работам, к текущему контролю успеваемости и подготовке к зачету, экзамену, курсовой работе.

После вводных лекций, в которых обозначается содержание дисциплины, ее проблематика и практическая значимость, студентам выдаются задания на лабораторные работы. Лабораторные работы охватывают модули 2-6.

В рамках дисциплины выполняется 20 лабораторных работ, которые защищаются устным опросом.

Выполнение всех лабораторных работ обязательно.

В случае невыполнения лабораторной работы по уважительной причине студент должен выполнить пропущенные лабораторные занятия в часы, отведенные на консультирование с преподавателем.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература по дисциплине

1. Воронов, В.И. Data Mining - технологии обработки больших данных : учебное пособие / В.И. Воронов, Л.И. Воронова, В.А. Усачев. - Москва : Московский технический университет связи и информатики, 2018. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: https://www.iprbookshop.ru/81324.html . - (ID=145723-0)

2. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных : учебное пособие / Н.Г. Ярушкина [и др.]. - Ульяновск : Ульяновский государственный технический университет, 2020. - ЦОР IPR SMART. - Текст : непосредственный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-9795-2088-9. - URL: https://www.iprbookshop.ru/106136.html . - (ID=145145-0)

3. Макшанов, А.В. Большие данные. Big Data : учебник / А.В. Макшанов, А.Е. Журавлев, Л.Н. Тындыкарь. - 2-е изд. - Санкт-Петербург [и др.] : Лань, 2022. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-9690-7. - URL: https://e.lanbook.com/book/198599 . - (ID=145911-0)

7.2. Дополнительная литература по дисциплине

1. Анализ данных : учебник для академического бакалавриата / В.С. Мхитарян [и др.]; под ред. В.С. Мхитаряна ; Национальный исследовательский ун-т - Высшая Школа Экономики. - Москва : Юрайт, 2022. - (Бакалавр. Академический курс). - Образовательная платформа Юрайт. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-534-00616-2. - URL: https://urait.ru/bcode/489100 . - (ID=100456-0)

2. Андрианова, Е.Е. Управление данными. Интеллектуальный анализ данных : учебное пособие / Е.Е. Андрианова, И.А. Липанова, О.Ю. Сабинин; Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича. - Санкт-Петербург : Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича, 2016. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: https://e.lanbook.com/book/180046 . - (ID=145914-0)

3. Вольфсон, М.Б. Анализ данных : учебное пособие / М.Б. Вольфсон; Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича. - Санкт-Петербург : Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича, 2015. - ЭБС Лань. -

Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: https://e.lanbook.com/book/180254 . - (ID=145915-0)

4. Головенчик, Г.Г. Цифровая экономика : учебно-методический комплекс для студентов, обучающихся по специальности «Мировая экономика» / Г.Г. Головенчик; Белорусский государственный университет. - Минск : Белорусский государственный университет, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-985-566-847-4. - URL: https://e.lanbook.com/book/180524 . - (ID=145905-0)

5. Железнов, М.М. Методы и технологии обработки больших данных : учебно-методическое пособие / М.М. Железнов. - Москва : МИСИ-МГСУ : ЭБС АСВ, 2020. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-7264-2193-3. - URL: https://www.iprbookshop.ru/101802.html . - (ID=145721-0)

6. Каленик, А.И. Использование новых возможностей Microsoft SQL Server 2005 / А.И. Каленик. - СПб. [и др.] ; М. : Питер : Русская редакция, 2006. - 321 с. : ил. - Текст : непосредственный. - ISBN 5-469-01404-5 (Питер) : 180 р. - (ID=66318-2)

7. Кобзаренко, Д. Н. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес» : учебное пособие / Д. Н. Кобзаренко, А. Г. Мустафаев ; составитель Д. Н. Кобзаренко. — Махачкала : ДГУНХ, 2019. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/246542 (дата обращения: 04.10.2022). — Режим доступа: для авториз. пользователей. - (ID=150643-0)

8. Методы и модели исследования сложных систем и обработки больших данных : монография / И.Ю. Парамонов [и др.]; Парамонов И.Ю., Смагин В.А., Косых Н.Е., Хомоненко А.Д. ; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург [и др.] : Лань, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-4006-1. - URL: https://e.lanbook.com/book/126938 . - (ID=136026-0)

9. Методы и модели исследования сложных систем и обработки больших данных : монография / И.Ю. Парамонов [и др.]; Парамонов И.Ю., Смагин В.А., Косых Н.Е., Хомоненко А.Д. ; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург [и др.] : Лань, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-4006-1. - URL: https://e.lanbook.com/book/126938 . - (ID=136026-0)

10. Миркин, Б.Г. Введение в анализ данных : учебник и практикум / Б.Г. Миркин. - Москва : Юрайт, 2022. - (Высшее образование). - Образовательная платформа Юрайт. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.09.2022. - ISBN 978-5-9916-5009-0. - URL: https://urait.ru/book/vvedenie-v-analiz-dannyh-469306 . - (ID=134346-0)

11. Новикова, О.А. Анализ данных : учебное пособие. Часть 1 / О.А. Новикова, Е.Г. Андрианова; МИРЭА - Российский технологический университет. - Москва : МИРЭА - Российский технологический университет, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: https://e.lanbook.com/book/167597 . - (ID=145912-0)

12. Просто о больших данных / Д. Гурвиц [и др.]. - Москва : Эксмо, 2015. - 394 с. - (Библиотека Сбербанка. Т.58). - Текст : непосредственный. - ISBN 978-5-699-85806-4 : 414 р. - (ID=114809-5)

13. Рындина, С.В. Цифровая трансформация бизнеса: использование аналитики на основе больших данных : учебное пособие / С.В. Рындина; Пензенский государственный университет. - Пенза : Пензенский государственный университет, 2019. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-907262-04-1. - URL: https://e.lanbook.com/book/162301 . - (ID=145921-0)

14. Синева, И.С. Анализ данных в среде R : учебное пособие. Ч. 1 / И.С. Синева. - Москва : Московский технический университет связи и информатики, 2018. - ЦОР IPR SMART. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - Текст : электронный. - URL: https://www.iprbookshop.ru/92422.html . - (ID=146948-0)

15. Федин, Ф.О. Анализ данных : учебное пособие. Ч. 2 : Инструменты Data Mining / Ф.О. Федин, Ф.Ф. Федин. - Москва : Московский городской педагогический университет, 2012. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: https://www.iprbookshop.ru/26445.html . - (ID=146080-0)

16. Целых, А.Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» / А.Н. Целых, А.А. Целых, Э.М. Котов. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-9275-3783-9. - URL: https://www.iprbookshop.ru/117165.htm . - (ID=145142-0)

7.3. Методические материалы

1. Вопросы по дисциплине "Управление и анализ данных", направление подготовки 09.04.01 Информатика и вычислительная техника, профиль - Информационное и программное обеспечение автоматизированных систем : в составе учебно-методического комплекса / Тверской гос. техн. ун-т, Каф. ЭВМ ; разраб. А.В. Попов. - Тверь : ТвГТУ, 2016. - (УМК-В). - Сервер. - Текст : электронный. - URL: https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/124790 . - (ID=124790-0)

2. Верпатова, О.Ю. Анализ данных в SPSS : учеб. пособие : в составе учебно-методического комплекса / О.Ю. Верпатова; Тверской гос. техн. ун-т. - Тверь : ТвГТУ, 2016. - 80 с. : ил. - (УМК-У). - Текст : непосредственный. - Текст : непосредственный. - ISBN 978-5-7995-0824-1 : [б. ц.]. - (ID=112484-75)

3. Методические указания по курсу "Компьютерный анализ данных в SPSS" для студентов направления подготовки бакалавров 040100 Социология и 080400 Управление персоналом дневной, заочной форм обучения и института заочного и дополнительного профессионального образования ТвГТУ / Тверской гос. техн. ун-т, Каф. СиСТ ; сост. А.В. Вайсбург. - Тверь : ТвГТУ, 2014. - 31 с. - Текст : непосредственный. - 33 р. 20 к. - (ID=102493-95)

4. Методические указания по курсу "Компьютерный анализ данных в SPSS" для студентов направления подготовки бакалавров 040100 Социология и 080400 Управление персоналом дневной, заочной форм обучения и института заочного и дополнительного профессионального образования ТвГТУ / Тверской гос. техн. ун-т, Каф. СиСТ ; сост. А.В. Вайсбург. - Тверь : ТвГТУ, 2014. - Сервер. - Текст : электронный. - 0-00. - URL: https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/102275 . - (ID=102275-1)

5. Комиссарчик, В.Ф. Анализ данных и планирование эксперимента : учеб. пособие : в составе учебно-методического комплекса / В.Ф. Комиссарчик; Тверской гос. техн. ун-т. - Тверь : ТвГТУ, 2000. - 135 с. - (УМК-У). - Текст : непосредственный. - 30 р. - (ID=5522-5)

7.4. Программное обеспечение по дисциплине

Операционная система Microsoft Windows: лицензии № ICM-176609 и № ICM-176613 (Azure Dev Tools for Teaching).
Microsoft Office 2007 Russian Academic: OPEN No Level: лицензия № 41902814.
WPS Office: MPL 1.1/GPL 2.0/LGPL 2.1.
Libre Office: MPL 2.0.
LMS Moodle: GPL 3.0.
Python 3.8.10

7.5. Специализированные базы данных, справочные системы, электронно-библиотечные системы, профессиональные порталы в Интернет

ЭБС и лицензионные ресурсы ТвГТУ размещены:

1. Ресурсы:https://lib.tstu.tver.ru/header/obr-res
2. ЭКТвГТУ:https://elib.tstu.tver.ru/MegaPro/Web
3. ЭБС "Лань":https://e.lanbook.com/
4. ЭБС "Университетская библиотека онлайн":https://www.biblioclub.ru/
5. ЭБС «IPRBooks»:https://www.iprbookshop.ru/
6. Электронная образовательная платформа "Юрайт" (ЭБС «Юрайт»):https://urait.ru/
7. Научная электронная библиотека eLIBRARY:https://elibrary.ru/
8. Информационная система "ТЕХНОРМАТИВ". Конфигурация "МАКСИМУМ" : сетевая версия (годовое обновление): [нормативно-технические, нормативно-правовые и руководящие документы (ГОСТы, РД, СНиПы и др.]. Диск 1,2,3,4. - М. :Технорматив, 2014. - (Документация для профессионалов). - CD. - Текст : электронный. - 119600 р. – (105501-1)
9. База данных учебно-методических комплексов:https://lib.tstu.tver.ru/header/umk.html

УМК размещен: https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/189526

8. Материально-техническое обеспечение

Кафедра «Электронных вычислительных машин» имеет аудитории для проведения лекций и лабораторных занятий по дисциплине; специализированные учебные классы, оснащенные современной компьютерной техникой, необходимым программным обеспечением, электронными учебными пособиями для проведения лабораторных работ и самостоятельной работы.

В наличии имеются презентационные мультимедийные лекционные курсы, разработанные преподавателями кафедры ЭВМ и внешними разработчиками.

Для проведения лабораторных работ имеются лаборатории с персональными компьютерами (наличие локальной вычислительной сети необязательно).

9. Оценочные средства для проведения промежуточной аттестации

9.1. Оценочные средства для проведения промежуточной аттестации в форме экзамена

1. Экзаменационный билет соответствует форме, утвержденной Положением о рабочих программах дисциплин, соответствующих федеральным государственным образовательным стандартам высшего образования с учетом профессиональных стандартов. Типовой образец экзаменационного билета приведен в Приложении. Обучающемуся даётся право выбора заданий из числа, содержащихся в билете, принимая во внимание оценку, на которую он претендует.

Число экзаменационных билетов – 10. Число вопросов (заданий) в экзаменационном билете – 3 (1 вопрос для категории «знать» и 2 вопроса для категории «уметь»).

Продолжительность экзамена – 60 минут.

2. Шкала оценивания промежуточной аттестации в форме экзамена – «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

3. Критерии оценки за экзамен:

для категории «знать»: выше базового – 2;

базовый – 1;

ниже базового – 0;

критерии оценки и ее значение для категории «уметь»:

отсутствие умения – 0 балл;

наличие умения – 2 балла.

«отлично» - при сумме баллов 5 или 6;

«хорошо» - при сумме баллов 4;

«удовлетворительно» - при сумме баллов 3;

«неудовлетворительно» - при сумме баллов 0, 1 или 2.

4. Вид экзамена – письменный экзамен, включающий решение задач с использованием ЭВМ.

5. База заданий, предъявляемая обучающимся на экзамене.

МОДУЛЬ 1 «Основные понятия машинного обучения»

1. Что такое искусственный интеллект? В чем связь между машинным обучением и искусственным интеллектом?
2. Объясните основные типы моделей машинного обучения (обучение с учителем, без учителя, с подкреплением). Приведите примеры.
3. Что такое обучающая выборка? Какие этапы включает в себя создание обучающей выборки?
4. Какие типы выборок существуют? Объясните назначение каждого типа и приведите примеры (например, выборка для обучения, тестовая выборка, валидационная выборка).
5. Какие проблемы могут возникнуть при создании обучающей выборки и как их можно решить? (например, несбалансированные данные, отсутствие данных)
6. Какие бывают типы данных для машинного обучения? (числовые, категориальные, текстовые и т.д.)
7. В чем заключаются основные этапы машинного обучения? (подготовка данных, выбор модели, обучение, оценка)

МОДУЛЬ 2 «Методы машинной кластеризации»

1. Что такое кластерный анализ? Объясните его цель и область применения.
2. Что такое геометрическая интерпретация кластерного анализа?
3. Опишите основные этапы кластерного анализа. Какие факторы влияют на каждый этап?
4. Какие методы снижения размерности признакового пространства вы знаете? Зачем они нужны?
5. Какие меры близости используются в кластерном анализе? Объясните их различия и когда их следует применять (например, евклидово расстояние, косинусное расстояние).
6. Какие существуют классификации методов кластерного анализа? Опишите основные типы.
7. Опишите эвристические алгоритмы кластеризации. Приведите примеры.
8. Опишите иерархический алгоритм кластеризации. Приведите примеры.
9. Опишите метод k-средних. Объясните его принцип работы и особенности.
10. Какие критерии используются для оценки качества кластеризации? Объясните каждый критерий.

МОДУЛЬ 3 «Методы машинной классификации»

1. Что такое классификация в машинном обучении? Объясните ее цель и область применения.
2. Какие ошибки могут возникать при классификации? Объясните различные типы ошибок и их последствия (ложноположительные, ложноотрицательные).
3. Опишите логические методы классификации. Приведите примеры.
4. Опишите Байесовский классификатор. Объясните его принцип работы и область применения.
5. Что такое деревья классификации? Объясните их принцип работы и преимущества.

6. Опишите алгоритмы построения деревьев классификации: CART и CHAID. В чем их различия?
7. Что такое Random Forest? Объясните его принцип работы и преимущества.

Пользование различными техническими устройствами, кроме ЭВМ компьютерного класса и программным обеспечением, необходимым для решения поставленных задач, не допускается. При желании студента покинуть пределы аудитории во время экзамена экзаменационный билет после его возвращения заменяется.

Преподаватель имеет право после проверки письменных ответов на экзаменационные вопросы и решенных на компьютере задач задавать студенту в устной форме уточняющие вопросы в рамках содержания экзаменационного билета, выданного студенту.

Иные нормы, регламентирующие процедуру проведения экзамена, представлены в Положении о текущем контроле успеваемости и промежуточной аттестации студентов.

9.2. Оценочные средства для проведения промежуточной аттестации в форме зачета

1. Шкала оценивания промежуточной аттестации – «зачтено», «не зачтено».

2. Вид промежуточной аттестации в форме зачета.

Промежуточная аттестация в форме зачета устанавливается преподавателем по результатам текущего контроля знаний и умений, обучающегося без дополнительных контрольных испытаний.

При промежуточной аттестации без выполнения дополнительного итогового контрольного испытания студенту в обязательном порядке описываются критерии проставления зачёта:

«зачтено» - выставляется обучающемуся при условии выполнения им всех контрольных мероприятий, посещения лекций и лабораторных занятий в объеме, соответствующем не менее чем 80% от количества часов, отведенного на контактную работу с преподавателем.

3. Для дополнительного итогового контрольного испытания студенту в обязательном порядке предоставляется:

база заданий, предназначенных для предъявления обучающемуся на дополнительном итоговом контрольном испытании (типовой образец задания приведен в Приложении);

методические материалы, определяющие процедуру проведения дополнительного итогового испытания и проставления зачёта.

Задание выполняется письменно и с использованием ЭВМ.

Перечень вопросов дополнительного итогового контрольного испытания:
МОДУЛЬ 3 «Методы машинной классификации»

1. Опишите метод логистической регрессии. Объясните ее принцип работы и область применения.
2. Сущность метода дискриминантного анализа. Объясните его принцип работы и область применения.
3. Как оценивается качество классификатора? Какие метрики используются? (точность, полнота, F1-мера и т.д.)
4. Что такое A/B тестирование? Как оно используется для оценки качества классификаторов?

МОДУЛЬ 4 «Нейросетевые технологии»

1. Что такое нейронная сеть? Опишите основные компоненты нейронной сети.
2. Опишите модель нейрона (перцептрон). Объясните его принцип работы.
3. Что такое однослойные нейронные сети? Какие задачи они могут решать?
4. Что такое многослойные нейронные сети? Как они работают?
5. Что такое глубокое обучение? В чем его отличие от классических методов машинного обучения?
6. Что такое свертка (convolutions)? Зачем она используется?
7. Опишите сверточные нейронные сети (CNN). Объясните их архитектуру и область применения.
8. Опишите рекуррентные нейронные сети (RNN). Объясните их архитектуру и область применения.

МОДУЛЬ 5 «Приложения методов машинного обучения»

1. Объясните, как методы машинного обучения используются для тематического моделирования. Приведите примеры.
2. Как методы машинного обучения используются для выявления трендов в данных?
3. Какие методы машинного обучения используются для поиска информации? Приведите примеры.
4. Какие методы машинного обучения используются для анализа информационных потоков?
5. Как методы машинного обучения применяются для анализа биоинформационных данных?
6. Как методы машинного обучения используются для рубрикации документов?
7. Что такое Reinforcement Learning (RL)? Объясните его принцип работы и область применения.
8. Приведите примеры практического применения методов машинного обучения в различных областях.

МОДУЛЬ 6 «Технологии работы с большими данными»

1. Что такое большие данные? Какие основные характеристики больших данных вы знаете?
2. Какие проблемы возникают при обработке больших данных?
3. Что такое Hadoop? Объясните его основные принципы.
4. Опишите архитектуру Hadoop. Какие основные компоненты входят в ее состав?
5. Что такое HDFS? Объясните его назначение.

6. Что такое MapReduce? Объясните принцип работы этой технологии.
7. Какие компоненты Hadoop вы знаете? Опишите их назначение.
8. Объясните процесс установки и конфигурации Hadoop.
9. Какие современные инструменты и технологии используются для обработки больших данных помимо Hadoop? (Spark, etc.)
10. Какие существуют способы оптимизации обработки больших данных?

Пользование различными техническими устройствами, кроме ЭВМ компьютерного класса и программным обеспечением, необходимым для решения поставленных задач, не допускается.

Преподаватель имеет право после проверки письменных ответов на экзаменационные вопросы задавать студенту в устной форме уточняющие вопросы в рамках содержания экзаменационного билета, выданного студенту.

Иные нормы, регламентирующие процедуру проведения экзамена, представлены в Положении о текущем контроле успеваемости и промежуточной аттестации студентов.

Число заданий для дополнительного итогового контрольного испытания - 10.

Число вопросов – 3 (2 вопроса для категории «знать» и 1 вопрос для категории «уметь»).

Продолжительность – 60 минут.

4. Критерии выполнения контрольного испытания и условия проставления зачёта:

для категории «знать» (бинарный критерий):

ниже базового - 0 балл;

базовый уровень – 1 балла;

критерии оценки и ее значение для категории «уметь» (бинарный критерий):

отсутствие умения – 0 балл;

наличие умения – 1 балла.

Критерии итоговой оценки за зачет:

«зачтено» - при сумме баллов 2 или 3;

«не зачтено» - при сумме баллов 0 или 1.

9.3. Оценочные средства для проведения промежуточной аттестации в форме курсовой работы

1. Шкала оценивания курсовой работы – «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

2. Тема курсовой работы: «Разработка аналитического программного обеспечения».

3. Критерии итоговой оценки за курсовую работу.

Таблица 4. Оцениваемые показатели для проведения промежуточной аттестации в форме курсовой работы

№ раздела	Наименование раздела	Баллы по шкале уровня
1	Термины и определения	Выше базового – 2 Базовый – 1 Ниже базового – 0
2	Введение	Выше базового – 2 Базовый – 1 Ниже базового – 0
3	Общая часть (обзор литературы по выбранной теме курсовой работы)	Выше базового – 6 Базовый – 3 Ниже базового – 0
4	Специальная часть	Выше базового – 10 Базовый – 6 Ниже базового – 0
5	Заключение	Выше базового – 2 Базовый – 1 Ниже базового – 0
6	Список использованных источников	Выше базового – 2 Базовый – 1 Ниже базового – 0

Критерии итоговой оценки за курсовую работу:

«отлично» – при сумме баллов от 20 до 24;

«хорошо» – при сумме баллов от 16 до 19;

«удовлетворительно» – при сумме баллов от 11 до 15;

«неудовлетворительно» – при сумме баллов менее 11, а также при любой другой сумме, если по разделу «Специальная часть», работа имеет 0 баллов.

4. В процессе выполнения курсовой работы руководитель осуществляет систематическое консультирование.

5. Дополнительные процедурные сведения:

- студенты выбирают тему для курсовой работы самостоятельно из предложенного списка и согласовывают свой выбор с преподавателем в течение двух первых недель обучения;

- проверку и оценку работы осуществляет руководитель, который доводит до сведения обучающего достоинства и недостатки курсовой работы, и ее оценку. Оценка проставляется в зачетную книжку обучающегося и ведомость для курсовой работы. Если обучающийся не согласен с оценкой руководителя, проводится защита работы перед комиссией, которую назначает заведующий кафедрой;

- защита курсовой работы проводится в течение двух последних недель семестра и выполняется в форме устной защиты в виде доклада и презентации на 5-7 минут с последующим ответом на поставленные вопросы, в ходе которых выясняется глубина знаний студента и самостоятельность выполнения работы;

- работа не подлежит обязательному внешнему рецензированию;

- курсовые работы хранятся на кафедре в течение трех лет.

10. Методические рекомендации по организации изучения дисциплины

Студенты перед началом изучения дисциплины ознакомлены с системами кредитных единиц и балльно-рейтинговой оценки.

Студенты, изучающие дисциплину, обеспечиваются электронными изданиями или доступом к ним, учебно-методическим комплексом по дисциплине, включая методические указания к выполнению лабораторных, и всех видов самостоятельной работы.

В учебный процесс рекомендуется внедрение субъект-субъектной педагогической технологии, при которой в расписании каждого преподавателя определяется время консультаций студентов по закрепленному за ним модулю дисциплины.

11. Внесение изменений и дополнений в рабочую программу дисциплины

Содержание рабочих программ дисциплин ежегодно обновляется протоколами заседаний кафедры по утвержденной «Положением о структуре, содержании и оформлении рабочих программ дисциплин по образовательным программам, соответствующим ФГОС ВО с учетом профессиональных стандартов» форме.

Приложение 1

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»

Направление подготовки бакалавров – 09.03.01. Информатика и вычислительная техника

Направленность (профиль) – Вычислительные машины, комплексы, системы и сети

Кафедра электронных вычислительных машин

Дисциплина «Машинное обучение и большие данные»

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Вопрос для проверки уровня «ЗНАТЬ» - 0 или 1, или 2 балла
Основные типы моделей машинного обучения (обучение с учителем, без учителя, с подкреплением). Приведите примеры.

2. Задание для проверки уровня «УМЕТЬ» - 0 или 2 балла
В чем заключаются основные этапы машинного обучения? (подготовка данных, выбор модели, обучение, оценка)

3. Задание для проверки уровня «УМЕТЬ» - 0 или 2 балла
Написать программу для считывания данных из csv файла и расчета показателей описательной статистики предложенной величины.

Критерии итоговой оценки за экзамен:
«отлично» - при сумме баллов 5 или 6;
«хорошо» - при сумме баллов 4;
«удовлетворительно» - при сумме баллов 3;
«неудовлетворительно» - при сумме баллов 0, 1 или 2 балла;

Составитель: к.т.н, доцент _____Е.Е. Фомина

Заведующий кафедрой: к.т.н., профессор _____А.Р. Хабаров

Приложение 2

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»

Направление подготовки бакалавров – 09.03.01. Информатика и вычислительная техника

Направленность (профиль) – Вычислительные машины, комплексы, системы и сети

Кафедра электронных вычислительных машин

Дисциплина «Машинное обучение и большие данные»

ЗАДАНИЕ ДЛЯ ДОПОЛНИТЕЛЬНОГО ИТОГОВОГО КОНТРОЛЬНОГО ИСПЫТАНИЯ № 1

1. Вопрос для проверки уровня «ЗНАТЬ» – 0 или 1 балл:
Дайте определение понятию Большие данные

2. Вопрос для проверки уровня «ЗНАТЬ» – 0 или 1 балл:
Алгоритм иерархического кластерного анализа.

3. Задание для проверки уровня «УМЕТЬ» – 0 или 1 балл:
Разработать модель логистической регрессии для решения задачи.

Критерии итоговой оценки за зачет:
«зачтено» - при сумме баллов 2 или 3;
«не зачтено» - при сумме баллов 0, или 1.

Составитель: к.т.н, доцент _____Е.Е. Фомина

Заведующий кафедрой: к.т.н., профессор _____А.Р. Хабаров