

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

УТВЕРЖДАЮ
Проректор по учебной работе
_____ М.А. Смирнов
«__» _____ 2025 г.

РАБОЧАЯ ПРОГРАММА

дисциплины, части формируемой участниками образовательных
отношений Блока 1 «Дисциплины (модули)»
«Методы и алгоритмы машинного обучения»

Направление подготовки бакалавров 09.03.04 Программная инженерия.

Направленность (профиль) — Разработка систем искусственного интеллекта.

Типы задач профессиональной деятельности — производственно-технологический.

Форма обучения – очная.

Факультет информационных технологий.

Кафедра «Программное обеспечение».

Тверь 2025

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

Разработчик программы

Е.И. Корнеева

Программа рассмотрена и одобрена на заседании кафедры ПО
« ____ » _____ 2025 г., протокол № _____.

Заведующий кафедрой ПО

А. Л. Калабин

Согласовано
Начальник УМО

Е.Э. Наумова

Начальник отдела
комплектования
зональной научной библиотеки

О. Ф. Жмыхова

1. Цель и задачи дисциплины

Целью изучения дисциплины «Методы и алгоритмы машинного обучения» является формирование у обучающихся представлений о методах и алгоритмах машинного обучения, уяснение основных принципов разработки, внедрения и применения систем машинного обучения для решения прикладных задач.

Задачами дисциплины являются:

- формирование представлений у обучающегося об основных понятиях, используемых в предметной области машинного обучения;
- формирование представлений об областях практического использования методов машинного обучения и проблемах, связанных с их внедрением;
- получение знаний о принципах построения моделей машинного обучения;
- получение практических навыков разработки моделей машинного обучения с использованием библиотек `scikit-learn`, `NumPy`, `pandas`.

2. Место дисциплины в структуре ОП

Дисциплина относится к части, формируемой участниками образовательных отношений, Блока 1 ОП ВО, определяет подготовку бакалавров по направлению Программная инженерия в использовании методов машинного обучения в дальнейшей учебной, научной и профессиональной деятельности. Она требует знания основ математического анализа, линейной алгебры, теории вероятностей и математической статистики.

В результате изучения данной дисциплины студенты будут знать основные методы и алгоритмы машинного обучения, уметь применять их для решения практических задач с помощью средств современных библиотек, с учётом прикладной специфики.

3. Планируемые результаты обучения по дисциплине

3.1. Перечень компетенций, закреплённых за дисциплиной в ОХОП

Компетенция, закреплённая за дисциплиной в ОХОП:

ПК-5. *Способен разрабатывать, адаптировать, применять в профессиональной деятельности алгоритмы, программные средства, системы и комплексы обработки данных, методы и алгоритмы машинного обучения, программно-технические платформы, электронные библиотеки, программные оболочки приложений, сетевые технологии для решения задач в сфере искусственного интеллекта и смежных областях.*

ПК-6. *Способен выбирать, применять и проводить экспериментальную проверку работоспособности программных компонентов систем, включающих модули по созданию искусственного интеллекта.*

Индикаторы компетенции, закрепленные за дисциплиной в ОХОП

ИПК-5.3. Умеет работать с библиотеками и программными интерфейсами систем глубокого обучения и нейронных сетей.

ИПК-5.4. Осуществляет сбор данных по запросам при помощи специализированных программ, форматирование, создание набора данных для машинного обучения, интеграция технологий машинного обучения в оптимизацию приложений для улучшения производительности и функциональности.

ИПК-5.9. Использует вычислительные методы и алгоритмы для анализа и извлечения значимой информации из больших объемов различного контента с применением различных методов обработки, алгоритмов машинного обучения, интеллектуального анализа данных.

ИПК-6.1. Выбирает, комбинирует и адаптирует существующие программные продукты, для решения необходимых функций, профессиональных задач предприятий или организаций.

ИПК-6.2. Самостоятельно создает на основе стандартных оболочек с привлечением искусственного интеллекта программное обеспечение для решения необходимых функций, профессиональных задач предприятий или организаций.

Показатели оценивания индикаторов достижения компетенций

Знать:

- З1. основные парадигмы машинного обучения;
- З2. классические алгоритмы машинного обучения;
- З3. методы оценки качества моделей;
- З4. методы предобработки данных и отбора признаков.

Уметь:

- У1. формализовывать прикладные задачи в терминах машинного обучения;
- У2. реализовывать алгоритмы машинного обучения с использованием библиотек Python;

Иметь опыт практической подготовки:

- ПП1. оценивать эффективность подходов, использованных в разработке моделей.

3.2. Технологии, обеспечивающие формирование компетенций

Проведение лекционных занятий, выполнение лабораторных работ, самостоятельная работа под руководством преподавателя, выполнение курсовой работы.

4. Трудоемкость дисциплины и виды учебной работы

Таблица 1. Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачётных единиц	Академических часов
Общая трудоемкость дисциплины	7	252
Аудиторные занятия (всего)		120
В том числе:		
Лекции		60
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		60
Самостоятельная работа (всего)		96+36(экз)
В том числе:		
Курсовая работа (КР)		45
Курсовой проект (КП)		не предусмотрен
Расчетно-графические работы		не предусмотрены
Реферат		не предусмотрен
Другие виды самостоятельной работы: - подготовка к защите лабораторных работ		48
Контроль текущий и промежуточный (балльно-рейтинговый, экзамен)		36(экз)
Контроль текущий и промежуточный (балльно-рейтинговый, зачёт)		3
Практическая подготовка при реализации дисциплины (всего)		105
Практические занятия (ПЗ)		не предусмотрены

Лабораторные работы (ЛР)		60
Курсовая работа (КР)		45
Курсовой проект (КП)		не предусмотрен

5. Структура и содержание дисциплины

Структура и содержание дисциплины построены по модульно-блочному принципу. Под модулем дисциплины понимается укрупненная логико-понятийная тема, характеризующаяся общностью использованного понятийно-терминологического аппарата.

5.1. Структура дисциплины

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

№	Наименование модуля	Труд-ть (часы)	Лекции	Практич. занятия	Лаб. работы	Сам. работа
1	Введение в машинное обучение. Линейные модели	84	20	-	20	32+12(экз)
2	Непараметрические методы и деревья решений	84	20	-	20	32+12(экз)
3	Ансамблевые методы и продвинутое техники	84	20	-	20	32+12(экз)
	Всего на дисциплину	252	60	-	60	96+36(экз)

5.2. Содержание дисциплины

МОДУЛЬ 1: «ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ. ЛИНЕЙНЫЕ МОДЕЛИ»

Основы машинного обучения. Определение машинного обучения. История развития. Типы задач машинного обучения: классификация, регрессия, кластеризация. Парадигмы обучения: обучение с учителем,

обучение без учителя, обучение с подкреплением, полуконтролируемое обучение. Основные этапы решения задач машинного обучения. Понятие обучающей, валидационной и тестовой выборок. Базовые концепции: пространство признаков, функция потерь, целевая функция.

Линейная регрессия. Математическая постановка задачи линейной регрессии. Метод наименьших квадратов (МНК). Аналитическое решение задачи линейной регрессии. Градиентный спуск и его модификации (стохастический градиентный спуск, mini-batch). Полиномиальная регрессия и проблема переобучения. Регуляризация: Ridge (L2), Lasso (L1), ElasticNet. Метрики качества регрессии: MSE, RMSE, MAE, R^2 .

Линейная классификация. Постановка задачи бинарной классификации. Логистическая регрессия: функция активации, логистическая функция потерь. Многоклассовая классификация: one-vs-rest, one-vs-one, softmax регрессия. Методы оптимизации для классификации. Регуляризация в задачах классификации. Метрики качества классификации: accuracy, precision, recall, F1-score. ROC-кривая и AUC-ROC. Матрица ошибок (confusion matrix).

Предобработка данных и работа с признаками. Типы данных и их особенности. Обработка пропущенных значений. Масштабирование признаков: нормализация, стандартизация. Кодирование категориальных признаков: one-hot encoding, label encoding, target encoding. Отбор признаков: фильтрация, wrapper methods, embedded methods. Конструирование признаков (feature engineering). Работа с несбалансированными данными.

МОДУЛЬ 2: «НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ И ДЕРЕВЬЯ РЕШЕНИЙ»

Метод ближайших соседей (k-NN). Принцип работы алгоритма k-NN. Метрики расстояния: евклидова, манхэттенская, Минковского, косинусная. Выбор оптимального значения k. Взвешенный k-NN. Ускорение поиска ближайших соседей: KD-деревья, Ball-trees. Применение k-NN для классификации и регрессии. Преимущества и недостатки метода.

Метод опорных векторов (SVM). Разделяющая гиперплоскость и концепция максимального разделяющего зазора. Жесткий SVM для линейно разделимых данных. Мягкий SVM для линейно неразделимых данных. Функции ядра (kernel trick): линейное, полиномиальное, RBF, сигмоидное. Параметры SVM: C, gamma. SVM для регрессии (SVR). Многоклассовый SVM.

Деревья решений. Принцип построения деревьев решений. Критерии разбиения: энтропия, Gini, дисперсия. Алгоритмы построения деревьев: ID3, C4.5, CART. Проблема переобучения в деревьях решений.

Методы предотвращения переобучения: ограничение глубины, минимальное количество объектов в листе. Отсечение (pruning): pre-pruning, post-pruning. Деревья решений для регрессии. Интерпретируемость и важность признаков.

Методы оценки качества моделей. Разделение данных: hold-out, кросс-валидация. K-fold cross-validation, stratified k-fold, leave-one-out. Кривые обучения (learning curves). Кривые валидации (validation curves). Переобучение и недообучение: диагностика и методы борьбы. Bias-variance tradeoff (компромисс между смещением и разбросом). Подбор гиперпараметров: grid search, random search, Bayesian optimization.

МОДУЛЬ 3: «Ансамблевые методы и продвинутые техники»

Ансамблевые методы: бэггинг. Концепция ансамблевого обучения. Bootstrap aggregating (бэггинг). Случайный лес (Random Forest). Параметры случайного леса: количество деревьев, глубина, количество признаков. Оценка out-of-bag (OOB). Важность признаков в случайном лесу. Extra Trees. Преимущества и применение бэггинга.

Ансамблевые методы: бустинг. Принцип градиентного бустинга. AdaBoost: адаптивный бустинг. Gradient Boosting Machines (GBM). Современные реализации бустинга: XGBoost (eXtreme Gradient Boosting), LightGBM (Light Gradient Boosting Machine), CatBoost (Categorical Boosting). Параметры и регуляризация в бустинге. Особенности работы с категориальными признаками. Сравнение различных методов бустинга.

Стекинг и другие методы ансамблирования. Stacking (стекинг, суперобучение). Blending. Voting: hard voting, soft voting. Каскадные модели. Выбор базовых моделей для ансамбля. Практические рекомендации по построению ансамблей.

Обучение без учителя: кластеризация. Постановка задачи кластеризации. K-means алгоритм: принцип работы, выбор k, инициализация центроидов. Иерархическая кластеризация: агломеративная и дивизивная. DBSCAN: кластеризация на основе плотности. Метрики качества кластеризации: силуэт, индекс Дэвиса-Болдина. Применение кластеризации: сегментация, сжатие данных, обнаружение аномалий.

Снижение размерности. Проблема размерности (curse of dimensionality). Метод главных компонент (PCA). t-SNE для визуализации многомерных данных. UMAP (Uniform Manifold Approximation and Projection). Автокодировщики для снижения размерности. Применение методов снижения размерности.

Специальные темы. Работа с временными рядами. Обнаружение аномалий (anomaly detection). Системы рекомендаций: коллаборативная фильтрация, content-based методы. Transfer learning и domain adaptation.

Интерпретируемость моделей машинного обучения. Этические аспекты и bias в машинном обучении. MLOps: жизненный цикл моделей машинного обучения в production.

5.3. Лабораторные работы

Таблица 3. Лабораторные работы и их трудоемкость

Модули. Цели лабораторных работ	Примерная тематика лабораторных работ	Трудоемкость в часах
Модуль 1 Цель: освоение базовых алгоритмов машинного обучения, предобработки данных и оценки моделей	Линейная регрессия и методы оптимизации	3
	Логистическая регрессия и метрики качества классификации	5
	Предобработка данных и feature engineering	2
	Регуляризация и кросс-валидация	6
	Анализ данных и визуализация	4
Модуль 2 Цель: изучение непараметрических методов и деревьев решений, освоение техник валидации и подбора гиперпараметров	Метод k-NN и метрики расстояния	4
	Метод опорных векторов и функции ядра	3
	Деревья решений и pruning	5
	Кросс-валидация и подбор гиперпараметров	4
	Диагностика переобучения и недообучения	4
Модуль 3 Цель: освоение ансамблевых методов, методов кластеризации и снижения размерности	Случайный лес и важность признаков	4
	Градиентный бустинг (XGBoost, LightGBM)	4
	Сравнение ансамблевых методов	4

	Кластеризация (K-means, DBSCAN)	4
	Снижение размерности (PCA, t-SNE)	4

5.4. Практические 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 и содержание 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, экзамену, зачёту с оценкой, в выполнении курсовой работы.

В рамках дисциплины выполняется 15 лабораторных работ (5-й семестр — 5 работ, 6-й семестр — 10 работ), которые защищаются посредством тестирования или устным опросом (по желанию обучающегося). Максимальная оценка за каждую выполненную лабораторную работу – 5 баллов, минимальная – 3 балла.

Выполнение всех лабораторных работ обязательно. В случае невыполнения лабораторной работы по уважительной причине студент имеет право выполнить письменный реферат, по согласованной с преподавателем теме по модулю, по которому пропущена лабораторная работа. Возможная тематическая направленность реферативной работы для каждого учебно-образовательного модуля представлена в следующей таблице:

Таблица 4. Темы рефератов

№ п/п	Модули	Возможная тематика самостоятельной реферативной работы
1.	Модуль 1	Сравнительный анализ методов регуляризации. Современные методы оптимизации в машинном

		обучении. Байесовский подход к машинному обучению.
2.	Модуль 2	Методы ускорения алгоритма k-NN. Kernel methods в машинном обучении. Продвинутые техники построения деревьев решений.
3.	Модуль 3	Сравнительный анализ библиотек градиентного бустинга. AutoML: автоматизация машинного обучения. Интерпретируемость моделей машинного обучения (SHAP, LIME).

Оценивание в этом случае осуществляется путем устного опроса, а также по содержанию и качеству выполненного реферата.

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература по дисциплине

1. Горюшкин, А.П. Математическая логика и теория алгоритмов : учебник : в составе учебно-методического комплекса / А.П. Горюшкин. - Саратов : Вузовское образование, 2022. - (УМК-У). - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ЦОР IPR SMART. - ISBN 978-5-4487-0808-4. - URL: <https://www.iprbookshop.ru/117296.html> . - (ID=144996-0)
2. Крупский, В. Н. Теория алгоритмов. Введение в сложность вычислений : учебник для вузов / В. Н. Крупский. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 91 с. — (Высшее образование). — ISBN 978-5-534-21288-4. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/563580> (дата обращения: 17.12.2025). - (ID=142651-0).

7.2. Дополнительная литература по дисциплине

1. Вайнштейн, Ю. В. Математическая логика и теория алгоритмов : учебное пособие / Ю. В. Вайнштейн, Т. Г. Пенькова, В. И. Вайнштейн. — Красноярск : СФУ, 2019. — 110 с. — ISBN 978-5-7638-4076-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/157585> . - (ID=145006-0)
2. Виноградов, Г.П. Теория алгоритмов и элементы теории формальных языков : учеб. пособие для студентов направлений подготовки

бакалавра 15.03.04 Автоматизация технол. процессов и производств профиля "Технология и автоматизация производства в машиностроении" и 09.03.02 Информ. системы и технологии : в составе учебно-методического комплекса / Г.П. Виноградов, В.Н. Богатилов; Тверской государственный технический университет, Кафедра ИПМ. - Тверь :ТвГТУ, 2016. - (УМК-У). - Сервер. - Текст : электронный. - ISBN 978-5-7995-0845-6 : 0-00. - URL: <http://elibr.tstu.tver.ru/MegaPro/GetDoc/Megapro/113354> . - (ID=113354-1)

3. Гамова, А.Н. Математическая логика и теория алгоритмов : учебное пособие / А.Н. Гамова. - 4-е изд., доп. - Саратов : Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, 2020. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.09.2022. - ЭБС Лань. - ISBN 978-5-292-04649-3. - URL: <https://e.lanbook.com/book/170590> . - (ID=111573-0)
4. Глухов, М.М. Математическая логика. Дискретные функции. Теория алгоритмов : учебное пособие для вузов по спец. и напр. по информ. безопасности / М.М. Глухов, А.Б. Шишков. - Санкт-Петербург [и др.] : Лань, 2012. - 405 с. - (Учебники для вузов. Специальная литература). - Текст : непосредственный. - ISBN 978-5-8114-1344-7 : 766 р. 92 к. - (ID=95689-3)
5. Задачи и упражнения по математической логике, дискретным функциям и теории алгоритмов : учебное пособие / М.М. Глухов [и др.]; Глухов М.М., Козлитин О.А., Шапошников В.А., Шишков А.Б. - Санкт-Петербург : Лань, 2022. - табл. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 01.12.2025. - ЭБС Лань. - ISBN 978-5-507-44852-4. - URL: <https://e.lanbook.com/book/247400> . - (ID=189488-0)
6. Гамова, А.Н. Математическая логика и теория алгоритмов : учебное пособие / А.Н. Гамова. - 4-е изд., доп. - Саратов : Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, 2020. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.09.2022. - ЭБС Лань. - ISBN 978-5-292-04649-3. - URL: <https://e.lanbook.com/book/170590> . - (ID=111573-0)
7. Лавров, И.А. Задачи по теории множеств, математической логике и теории алгоритмов : [учеб. пособие для вузов] / И.А. Лавров, Л.Л. Максимова. - 5-е изд. ;испр. - Москва :Физматлит, 2004. - 255 с. -

Библиогр. : с. 248 - 249. - Текст : непосредственный. - ISBN 5-9221-0026-2 : 134 р. 64 к. - (ID=22585-5)

8. Судоплатов, С.В. Математическая логика и теория алгоритмов : учебник и практикум для вузов : в составе учебно-методического комплекса / С.В. Судоплатов, Е.В. Овчинникова. - 5-е изд., стер. - Москва : Юрайт, 2025. - (Высшее образование) (УМК-У). - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 30.08.2022. - Образовательная платформа Юрайт. - ISBN 978-5-534-12274-9. - URL: <https://urait.ru/bcode/559978> . - (ID=142652-0)
9. Широков, Д. В. Теория алгоритмов : учебное пособие / Д. В. Широков. — Киров :ВятГУ, 2017. — 163 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/134610> . - (ID=111523-0).

7.3. Методические материалы

1. Учебно-методический комплекс дисциплины обязательной части Блока 1 "Методы и алгоритмы машинного обучения". Направление подготовки 09.03.04 Программная инженерия. Направленность (профиль) - Разработка систем искусственного интеллекта : ФГОС 3++ / Каф. Программное обеспечение ; сост. Е.И. Корнеева. - 2025. - (УМК). - Текст : электронный. - URL: <https://elibr.tstu.tver.ru/MegaPro/GetDoc/Megapro/189499> . - (ID=189499-0)

7.4. Программное обеспечение по дисциплине

- ОС "Альт Образование" 8
- Учебный комплект программного обеспечения КОМПАС-3D v18 для преподавателя
- Программное обеспечение КОМПАС-3D v18
- МойОфис Стандартный
- WPS Office
- Libre Office
- Lotus Notes!Domino,
- LMS Moodle
- Marc-SQL
- МегаПро,
- Office для дома и учебы 2013
- 7zip,
- ОС РЕД ОС
- ПО PIX.

- Python (версия 3.x): основные библиотеки для работы с данными и машинного обучения — pandas, numpy, scipy, scikit-learn, tensorflow, keras.
- Jupyter Notebooks: интерактивная среда для разработки, тестирования и визуализации алгоритмов машинного обучения.
- Средства визуализации данных: Matplotlib, Seaborn, Plotly — для построения графиков, диаграмм и интерактивных визуализаций результатов моделей.
- Инструменты для работы с данными: базы данных PostgreSQL и MySQL для хранения и выборки обучающих данных.
- Инструменты для параллельной обработки и масштабирования данных: Apache Spark с Python API (PySpark) для распределенного обучения моделей и анализа больших наборов данных.

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. ЭБ ТвГТУ: <https://elib.tstu.tver.ru/MegaPro/Web>
3. ЭБС "Лань": <https://e.lanbook.com/>
4. ЭБС "Университетская библиотека онлайн": <https://biblioclub.ru/>
5. Национальная электронная библиотека: <https://rusneb.ru>
6. ЦОР IPRSmart: <https://www.iprbookshop.ru/>
7. Электронная образовательная платформа "Юрайт": <https://urait.ru/>
8. Научная электронная библиотека eLIBRARY: <https://elibrary.ru/>
9. 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)
10. База данных учебно-методических комплексов: <https://lib.tstu.tver.ru/header/umk.html>

Профессиональные ресурсы:

10. Machine Learning на сайте К.В. Воронцова: <http://www.machinelearning.ru/>
11. Kaggle: <https://www.kaggle.com/>
12. Papers with Code: <https://paperswithcode.com/>
13. ArXiv.org: <https://arxiv.org/list/cs.LG/recent>

14. Scikit-learn Documentation: <https://scikit-learn.org/>

15. UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/>

УМК размещен: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/189499>

8. Материально-техническое обеспечение

При изучении дисциплины «Методы и алгоритмы машинного обучения» используются современные средства обучения: наглядные пособия, диаграммы, схемы.

Возможна демонстрация лекционного материала с помощью оверхед-проектора (кодоскопа) и мультипроектора.

Вуз имеет лабораторию для реализации лабораторного практикума по дисциплине «Методы и алгоритмы машинного обучения», учебный класс для проведения самостоятельной работы по курсу, оснащённый современной компьютерной и офисной техникой, необходимым программным обеспечением, электронными учебными пособиями, имеющий безлимитный выход в глобальную сеть; аудиторию для проведения семинарских занятий, практикумов и презентаций студенческих работ, оснащённую аудиовизуальной техникой.

9. Оценочные средства для проведения промежуточной аттестации

9.1. Оценочные средства для проведения промежуточной аттестации в форме экзамена

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

2. 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 баллов;

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.

3. Вид экзамена – письменный экзамен.

4. 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 – 20. Число вопросов (заданий) в экзаменационном билете – 3.

Продолжительность экзамена – 60 минут.

5. База заданий, предъявляемая обучающимся на экзамене.

1. Основные парадигмы машинного обучения. Обучение с учителем, без учителя, с подкреплением.
2. Линейная регрессия: математическая постановка, метод наименьших квадратов, градиентный спуск.
3. Логистическая регрессия для бинарной классификации. Функция потерь и оптимизация.
4. Метрики качества классификации: accuracy, precision, recall, F1-score. ROC-AUC.
5. Проблема переобучения и методы регуляризации: Ridge, Lasso, ElasticNet.
6. Bias-variance tradeoff. Компромисс между смещением и разбросом.
7. Методы кросс-валидации: k-fold, stratified k-fold, leave-one-out.
8. Метод k ближайших соседей. Выбор метрики расстояния и параметра k.
9. Метод опорных векторов (SVM). Концепция максимального разделяющего зазора.
10. Функции ядра в SVM: линейное, полиномиальное, RBF ядро.
11. Деревья решений: критерии разбиения (энтропия, Gini, дисперсия).
12. Проблема переобучения в деревьях решений. Методы отсечения (pruning).
13. Бэггинг и случайный лес (Random Forest). Принцип работы и преимущества.
14. Градиентный бустинг: принцип работы, отличия от бэггинга.
15. Сравнение XGBoost, LightGBM и CatBoost.
16. Методы ансамблирования: стекинг, блендинг, voting.
17. Важность признаков в деревьях решений и случайном лесе.
18. Методы отбора признаков: filter, wrapper, embedded методы.
19. Предобработка данных: масштабирование, нормализация, кодирование категорий.
20. Работа с несбалансированными классами: техники балансировки.
21. K-means алгоритм кластеризации. Выбор количества кластеров.
22. Иерархическая кластеризация: агломеративная и дивизивная.
23. DBSCAN: кластеризация на основе плотности.
24. Метод главных компонент (PCA). Снижение размерности.

25. Метрики качества регрессии: MSE, RMSE, MAE, R^2 .
26. Полиномиальная регрессия и выбор степени полинома.
27. Многоклассовая классификация: one-vs-rest, one-vs-one, softmax.
28. Обнаружение выбросов и методы их обработки.
29. Feature engineering: конструирование признаков.
30. Grid search и random search для подбора гиперпараметров.
31. Кривые обучения (learning curves) и их интерпретация.
32. Методы работы с пропущенными значениями.
33. Матрица ошибок (confusion matrix) и её интерпретация.
34. Стохастический градиентный спуск и его модификации.
35. Адаптивные методы оптимизации: Adam, RMSprop, AdaGrad.
36. Batch normalization в контексте обучения моделей.
37. Early stopping как метод регуляризации.
38. Проклятие размерности (curse of dimensionality).
39. Методы визуализации многомерных данных: t-SNE, UMAP.
40. Интерпретируемость моделей машинного обучения: SHAP, LIME.

Практические задачи:

1. Реализовать линейную регрессию методом градиентного спуска на Python.
2. Вычислить метрики качества классификации по заданной матрице ошибок.
3. Написать код для нормализации и стандартизации признаков.
4. Реализовать k-fold кросс-валидацию для заданной модели.
5. Построить ROC-кривую и вычислить AUC для бинарного классификатора.
6. Реализовать алгоритм k-means на Python.
7. Написать функцию для вычисления энтропии и Gini index.
8. Применить PCA для снижения размерности данных.
9. Реализовать one-hot encoding для категориальных признаков.
10. Написать код для обучения и оценки случайного леса на scikit-learn.
11. Выполнить подбор гиперпараметров с помощью GridSearchCV.
12. Реализовать функцию для вычисления расстояния между векторами.
13. Написать код для визуализации границ решения классификатора.
14. Применить методы балансировки классов (oversampling, undersampling).
15. Реализовать voting classifier на основе нескольких базовых моделей.
16. Написать код для анализа важности признаков в дереве решений.

17. Реализовать функцию для вычисления силуэта (silhouette score) кластеризации.
18. Применить регуляризацию L1 и L2 к логистической регрессии.
19. Написать код для построения кривой обучения (learning curve).
20. Реализовать функцию для обработки пропущенных значений.

При ответе на вопросы экзамена допускается использование справочными данными, ГОСТами, методическими указаниями по выполнению лабораторных работ в рамках данной дисциплины.

Пользование различными техническими устройствами 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 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 с оценкой

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

2. Вид промежуточной аттестации устанавливается преподавателем:

- по результатам текущего контроля знаний обучающегося без дополнительных контрольных испытаний:

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

«зачтено» — выставляется обучающемуся при условии выполнения им всех контрольных мероприятий: выполнения и защиты всех лабораторных работ.

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

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

2. Примерная тематика курсовой работы.

Каждому обучающемуся выдаётся индивидуальный вариант для разработки программного обеспечения. Студент по согласованию с преподавателем может самостоятельно выбрать тему для курсовой работы.

3. Критерии оценки качества выполнения как по отдельным разделам курсовой работы, так и работы в целом.

Разделы курсовой работы по дисциплине «Методы и алгоритмы машинного обучения»:

№ раздела	Наименование раздела	Баллы по шкале уровня
-	Нормативные ссылки	Выше базового — 2 Базовый — 1 Ниже базового — 0
-	Термины и определения	Выше базового — 2 Базовый — 1 Ниже базового — 0
-	Введение	Выше базового — 2 Базовый — 1 Ниже базового — 0
1	Общая часть (обзор литературы по теме курсовой работы, описание данных, постановка задачи)	Выше базового — 6 Базовый — 3 Ниже базового — 0
2	Специальная часть (описание алгоритмов, разработка и обучение моделей, эксперименты)	Выше базового — 6 Базовый — 3 Ниже базового — 0
-	Заключение	Выше базового — 2 Базовый — 1 Ниже базового — 0
-	Список использованных источников	Выше базового — 2 Базовый — 1 Ниже базового — 0
-	Приложения (исходный код, дополнительные графики и таблицы)	Выше базового — 6 Базовый — 3 Ниже базового — 0

Критерии итоговой оценки за курсовую работу:

- «отлично» – при сумме баллов от 24 до 28;
- «хорошо» – при сумме баллов от 19 до 23;
- «удовлетворительно» – при сумме баллов от 14 до 18;
- «неудовлетворительно» – при сумме баллов менее 14.

Методические материалы, определяющие процедуру выполнения и представления работы и технологию её оценивания.

Требования и методические указания по структуре, содержанию и выполнению работы, а также критерии оценки оформлены в качестве отдельно выпущенного документа.

Курсовая работа состоит из титульного листа, содержания, нормативных ссылок, терминов и определений, сокращений, введения, основной части, экспериментальной части, заключения, списка использованных источников и приложений. Текст должен быть структурирован, содержать рисунки и таблицы. Рисунки и таблицы должны располагаться сразу после ссылки на них в тексте таким образом, чтобы их можно было рассматривать без поворота курсовой работы. Если это сложно, то допускается поворот по часовой стрелке.

Если таблицу приходится переносить на следующую страницу, то помещают слова: «продолжение табл.» с указанием номера справа, графы таблицы пронумеровывают и повторяют их нумерацию на следующей странице. Заголовок таблицы не повторяют.

Во введении необходимо отразить актуальность темы исследования, цель и задачи курсовой работы. Объем должен составлять 1-2 страницы.

Теоретическая часть должна содержать обзор актуальных литературных и нормативных источников выбранного объекта курсовой работы.

В практической части необходимо отразить:

- проектирование структуры ПО;
- проектирование интерфейса ПО;
- ход реализации разрабатываемого ПО;
- примеры тестирования ПО;

В заключении необходимо раскрыть особенности отображения в курсовой работе поставленных задач. Объем должен составлять 1-2 страницы.

Список использованных источников должен содержать не менее 10 наименований (книг, журналов, электронных ресурсов и др.).

Дополнительные процедурные сведения:

а) Студенты выбирают тему для курсовой работы самостоятельно из предложенного списка и согласовывают свой выбор с преподавателем в течение первых двух недель обучения. К середине семестра на проверку представляется общая часть курсовой работы, за две недели до защиты – окончательный вариант.

б) проверку и оценку работы осуществляет руководитель, который доводит до сведения обучающегося достоинства и недостатки курсовой работы и ее оценку. Оценка проставляется в зачётную книжку обучающегося и ведомость для курсовой работы. Если обучающийся не

согласен с оценкой руководителя, проводится защита работы перед комиссией, которую назначает заведующий кафедрой;

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

г) работа не подлежит обязательному внешнему рецензированию.

В процессе выполнения обучающимся курсовой работы руководитель осуществляет систематическое консультирование.

Оптимальный объем курсовой работы 20-30 страниц машинописного текста (не включая приложения), набранного 12-14 шрифтом через 1.5 интервала на листах формата А4 с одной стороны. Поля должны составлять 20 мм сверху и снизу, 30 мм слева и 15 мм справа.

Источники использованной литературы должны оформляться согласно ГОСТ 7.1-2003 «Библиографическая запись. Библиографическое описание. Общие требования и правила составления». Список источников следует составлять в порядке упоминания их в тексте. Ссылки на источники должны приводиться по тексту в квадратных скобках.

Нумерация страниц курсовой работы должна быть сквозной. Первой страницей является титульный лист, на нем номер страницы не ставится, второй - содержание и т.д. Номер страницы проставляется арабскими цифрами снизу страницы, посередине. Приложения необходимо включать в сквозную нумерацию.

Защита курсовой работы проводится в течение двух последних недель семестра и выполняется в форме устной защиты в виде доклада 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.

В процессе выполнения обучающимся курсовой работы руководитель осуществляет систематическое консультирование.

9.4. Оценочные средства для проведения промежуточной аттестации в форме курсового проекта

Учебным планом курсовой проект по дисциплине не предусмотрен.

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, курсовых работ 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 закреплённому 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.

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Тверской государственный технический университет»

Направление подготовки бакалавров 09.03.04 Программная инженерия

Направленность (профиль) – Разработка систем искусственного
интеллекта.

Кафедра «Программное обеспечение»

Дисциплина «Методы и алгоритмы машинного обучения»

ЭКЗАМЕНАЦИОННЫЙ БИЛЕТ № 1

1. Вопрос для проверки уровня «ЗНАТЬ» – 0 или 1 или 2 балла:

Линейная регрессия: метод наименьших квадратов, градиентный спуск.

2. Задание для проверки уровня «УМЕТЬ» - 0 или 2 балла:

Дана матрица ошибок бинарного классификатора. Вычислите accuracy, precision, recall и F1-score. Интерпретируйте результаты.

3. Задание для проверки уровня «УМЕТЬ» – 0 или 2 балла:

Реализуйте алгоритм нормализации признаков (min-max scaling).
Объясните, когда следует применять нормализацию, а когда —
стандартизацию.

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.

Составитель: _____ Е.И. Корнеева

Заведующий кафедрой: д.ф.-м.н. _____ А.Л.Калабин