ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Приказом ректора
АНО ВО «Центральный университет»
от «02» июня 2026 г. № 0602.30

Рабочая программа дисциплины (модуля)
«Машинное обучение (Machine Learning)»
дополнительной профессиональной программы – программы профессиональной переподготовки «Машинное обучение в индустрии»

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2026

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Тематический план** 5
3. **Содержание дисциплины (модуля)** 5
4. **Учебно-методическое обеспечение** 7
5. **Материально-техническое обеспечение** 7
6. **Методические и оценочные материалы** 9

1. Краткая характеристика дисциплины (модуля)

Изучение дисциплины (модуля) «Машинное обучение (Machine Learning)» позволяет сформировать системное понимание методов анализа данных и разработки интеллектуальных моделей, применяемых для решения прикладных и бизнес-задач, а также научиться применять современные алгоритмы и инструменты машинного обучения для повышения качества принимаемых решений и создания конкурентоспособных цифровых продуктов.

Цель изучения дисциплины (модуля): формирование системного понимания методологии науки о данных (Data Science) и практических навыков разработки, валидации и интерпретации моделей машинного обучения для решения прикладных и бизнес-задач.

Задачи изучения дисциплины (модуля):

- освоить этапы жизненного цикла проекта машинного обучения и принципы работы ключевых алгоритмов машинного обучения для задач регрессии, классификации, кластеризации и uplift-моделирования;
- научиться проводить разведочный анализ данных, выполнять генерацию и отбор признаков, предотвращать утечки данных и готовить данные к моделированию;
- сформировать умение обучать, настраивать и валидировать модели с использованием кросс-валидации, пайплайнов, функций потерь и метрик качества;
- освоить методы борьбы с переобучением, регуляризации и интерпретации моделей с использованием SHAP, LIME и других инструментов объяснимого ИИ;
- развить навыки построения воспроизводимых решений с применением современных библиотек Python и интерпретации результатов в бизнес-контексте.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основы методологии науки о данных (Data Science), машинного обучения, включая этапы жизненного цикла проектов;
- принципы работы ключевых алгоритмов: линейные модели, метод k ближайших соседей (k-nearest neighbors, kNN), деревья решений, ансамбли, включая случайный лес (Random Forest) и бустинг (boosting), кластеризация и снижение размерности;
- метрики качества, функции потерь и методы валидации моделей для задач регрессии, классификации и uplift-моделирования;
- концепции смещения и разброса, переобучения, регуляризации, а также подходы к интерпретации моделей (SHAP, LIME);
- типы признаков и методы работы с ними, включая генерацию, отбор и предотвращение утечек данных.

уметь:

- проводить разведочный анализ данных (EDA), формулировать гипотезы и готовить данные для моделирования;
- обучать, настраивать и оценивать модели машинного обучения с использованием кросс-валидации, пайплайнов и метрик;

— реализовывать задачи кластеризации, детекции аномалий и снижения размерности, интерпретируя результаты;
— строить и валидировать uplift-модели, а также генерировать признаки для сложных данных (временных, географических, финансовых и других);
— интерпретировать модели и визуализировать влияние признаков с помощью SHAP, LIME и других инструментов.

владеть:

— инструментами Python для анализа и моделирования: pandas, numpy, scikit-learn;
— продвинутыми библиотеками машинного обучения: XGBoost, CatBoost, lightgbm, а также SHAP, LIME для интерпретации;
— навыками построения воспроизводимых пайплайнов и работы с разнотипными данными;
— методами предобработки, валидации и оптимизации моделей в реальных задачах.

2. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары (Практические занятия)			
1	Основные термины машинного обучения и смежных областей. Оценка качества моделей.	12	6		24	Домашние задания
2	Линейные модели	8	4	2	16	Домашние задания Коллоквиум
3	Деревья решений и модели, основанные на них	20	10	2	24	Домашние задания Коллоквиум
4	Обучение без учителя и работа с признаками	20	10	4	24	Домашние задания Проект
	Экзамен			***4***		
	Итого:	***60***	***30***	***12***	***88***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

3. Содержание дисциплины (модуля)

№ п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основные термины машинного обучения и смежных областей. Оценка качества моделей.	Жизненный цикл проекта машинного обучения (ML Lifecycle). Метрики и функционалы потерь в задаче регрессии. Метрики и функционалы потерь в задаче классификации. Контроль качества и выбор модели Метрические алгоритмы
2	Линейные модели	Линейная регрессия Логистическая регрессия, Метод опорных векторов (SVM)
3	Деревья решений и модели, основанные на них	Решающие деревья Сложность алгоритмов. Ансамбли алгоритмов Случайные леса Градиентный бустинг Аплифт-моделирование
4	Обучение без учителя и работа с признаками	Кластеризация и методы понижения размерности Обнаружение аномалий Генерация и отбор признаков

		Интерпретация моделей и диагностика сдвига данных Подведение итогов курса. Частые вопросы на собеседованиях по ML

4. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый слушатель в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Слушателям обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Бринк, Х. Машинное обучение : практическое руководство / Х. Бринк, Д. Ричардс, М. Феверолф. - Санкт-Петербург : Питер, 2018. - 336 с. - (Серия «Библиотека программиста»). - ISBN 978-5-496-02989-6. - Текст : электронный. - URL: https://znanium.com/catalog/product/1766396.

2. Лакшманан, В. Машинное обучение. Паттерны проектирования : практическое пособие / В. Лакшманан, С. Робинсон, М. Мунн. - Санкт-Петербург : БХВ-Петербург, 2022. - 448 с. - ISBN 978-5-9775-6797-8. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2140204.

Дополнительная литература:

1. Коэльо, Л. Построение систем машинного обучения на языке Python : практическое руководство / Л. Коэльо, В. Ричарт ; пер. с англ. А. А. Слинкина. - 3-е изд. - Москва : ДМК Пресс, 2023. - 304 с. - ISBN 978-5-89818-331-8. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102618.

2. Григорьев, А. Машинное обучение. Портфолио реальных проектов : практическое руководство / А. Григорьев. - Санкт-Петербург : Питер, 2023. - 496 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-1978-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/2123375.

5. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

— столами и стульями;

— компьютерной техникой;

— механическими калькуляторами;

— специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1	Катастрофы, стихийные бедствия, аварии, эпидемии. Солнечная и геомагнитная активность. /ежедневный обзор	http://www.disasters.chat.ru
2	Каталог по безопасности жизнедеятельности	http://www.eun.chat.ru
3	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
4	База данных для IT-специалистов	https://habr.com
5	База данных ScienceDirect	https://www.sciencedirect.com
6	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
7	Федеральный портал «Российское образование"	https://www.edu.ru/
8	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
9	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
10	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное

TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

6. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Машинное обучение (Machine Learning)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, коллоквиумы, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удается, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под

руководством преподавателя, где слушатели активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал, использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Коллоквиум – устные ответы на вопросы, список которых известен слушателю заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

Самостоятельная работа – работа слушателей, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы слушатели взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи слушателя включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***экзамена***.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Слушатель полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	
7	Хорошо	Слушатель обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	
5	Удовлетворительно	Слушатель обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Слушатель способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Слушатель не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Машинное обучение (Machine Learning)» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	10	Набор задач по темам недели
Коллоквиум	30%	3	Устные ответы на вопросы, список которых известен слушателю заранее
Проект	20%	1	Исследовательская работа по дисциплине (модулю) и презентация результатов
Экзамен	20%	1	Письменная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Машинное обучение (Machine Learning)»: «0,3 × среднее за домашние задания + 0,3 × среднее за коллоквиумы + 0,2 × проект + 0,2 × экзамен».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1

Выберите датасет для задачи регрессии или классификации, выполните разведочный анализ данных, разделите выборку на обучающую и тестовую, обучите не менее двух моделей (одну метрическую, например kNN), рассчитайте соответствующие метрики качества и функции потерь, выполните кросс-валидацию, сравните результаты моделей и обоснуйте выбор лучшей модели с точки зрения контроля качества и устойчивости.

Домашнее задание 2

Реализуйте линейную регрессию для задачи прогнозирования числового признака, оцените качество модели с использованием MAE, MSE и R^2, проанализируйте влияние регуляризации, затем решите задачу бинарной классификации с использованием логистической регрессии и SVM, сравните их по метрикам качества и визуализируйте границы принятия решений.

Домашнее задание 3

Обучите модель решающего дерева для задачи классификации или регрессии, проанализируйте глубину дерева и влияние гиперпараметров на переобучение, затем реализуйте случайный лес и градиентный бустинг (например, XGBoost или LightGBM), сравните модели по качеству и скорости обучения, выполните подбор гиперпараметров и сделайте выводы о преимуществах ансамблей по сравнению с одиночным деревом.

Домашнее задание 4

Выберите датасет и выполните кластеризацию (например, K-means или DBSCAN), примените метод снижения размерности (PCA или t-SNE) для визуализации, реализуйте алгоритм обнаружения аномалий, выполните генерацию и отбор признаков, проверьте наличие утечек данных, постройте uplift-модель для оценки инкрементального эффекта

(при наличии подходящих данных), интерпретируйте результаты с помощью SHAP или аналогичного инструмента и проанализируйте возможный сдвиг данных между обучающей и тестовой выборками.

Примерные вопросы к коллоквиуму

1. Что такое жизненный цикл проекта машинного обучения? Перечислите основные этапы и кратко охарактеризуйте каждый.
2. В чём разница между функционалом потерь и метрикой качества? Приведите примеры для задачи регрессии.
3. Как устроена логистическая регрессия? Как интерпретировать её коэффициенты?
4. Что такое переобучение? Какие методы позволяют с ним бороться?
5. В чём суть метода k ближайших соседей? Как выбор метрики влияет на качество модели?
6. Как устроено решающее дерево? Какие критерии используются для разбиения в задачах классификации?
7. Чем случайный лес отличается от градиентного бустинга? В каких случаях предпочтительнее каждый из методов?
8. Что такое кросс-валидация? Зачем она нужна и какие виды вы знаете?
9. Как работают методы понижения размерности, такие как PCA? В каких случаях их применение оправдано?
10. Что такое uplift-моделирование? В каких бизнес-задачах оно применяется?

Критерии оценивания коллоквиума:

1. Полнота и точность ответа — корректное раскрытие сути вопроса, отсутствие фактических ошибок.
2. Логичность и структурированность — чёткая последовательность изложения, наличие примеров и пояснений.
3. Использование терминологии — правильное употребление понятий и определений из области машинного обучения.
4. Связь с практикой — способность привести примеры применения методов в реальных задачах.
5. Ответ на дополнительные вопросы — умение корректно реагировать на уточнения и развёртывать ответ.

Примерное задание и критерии оценивания к проекту

Описание проекта:

Проект направлен на решение практической задачи машинного обучения (регрессии, классификации или аплифт-моделирования) с применением нескольких классов моделей, сравнением их качества и обоснованием выбора наилучшего решения.

Слушатель должен продемонстрировать владение основными понятиями машинного обучения, умение выбирать метрики и функционалы потерь, проводить контроль качества модели, работать с признаками и интерпретировать результаты.

В проекте необходимо реализовать и сравнить не менее трёх различных подходов:

- линейные модели;
- деревья решений и ансамбли;
- метрические алгоритмы;

• минимум один метод обучения без учителя (при обосновании применимости).

Цели проекта:

1. Закрепление понимания ключевых терминов и концепций машинного обучения.
2. Освоение методов оценки качества моделей и выбора метрик.
3. Практическое применение линейных моделей, деревьев решений, ансамблей и метрических алгоритмов.
4. Развитие навыков работы с признаками и интерпретации моделей.
5. Формирование навыков выбора и обоснования финальной модели.

Задачи проекта:

1. Постановка задачи

• Определить тип задачи и целевую переменную.

• Обосновать выбор метрик качества.

• Описать данные и их особенности.

2. Анализ и подготовка данных

• Провести исследовательский анализ данных.

• Обработать пропуски, выбросы и категориальные признаки.

• Выполнить генерацию и/или отбор признаков.

• При необходимости применить методы понижения размерности.

3. Обучение и сравнение моделей

Реализовать и сравнить:

• Линейные модели (линейная или логистическая регрессия, SVM).

• Деревья решений и ансамбли (решающее дерево, случайный лес, градиентный бустинг).

• Метрические алгоритмы (kNN).

• Минимум один метод обучения без учителя (кластеризация, обнаружение аномалий и др.).

4. Оценка качества и выбор модели

• Разделить данные на обучающую и тестовую выборки.

• Использовать кросс-валидацию.

• Проанализировать переобучение.

• Сравнить модели по выбранным метрикам.

• Обосновать выбор финальной модели.

5. Интерпретация и диагностика

• Проанализировать важность признаков.

• Провести анализ ошибок.

• Дать интерпретацию полученных результатов.

• Оценить устойчивость модели и возможный сдвиг данных (при наличии).

Критерии оценивания:

Оценка проекта осуществляется по следующим критериям:

• Корректность постановки задачи и использование терминологии машинного обучения.

• Обоснованность выбора метрик качества и функционалов потерь.

• Полнота и глубина анализа данных.

• Качество предобработки и работы с признаками.

• Реализация требуемых классов моделей.

• Корректность процедуры обучения и валидации.

- Сравнительный анализ моделей.
- Обоснованность выбора итоговой модели.
- Анализ переобучения и устойчивости модели.
- Интерпретация результатов и качество выводов.