Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»

Школа вычислительных социальных наук

УТВЕРЖДАЮ:
Ректор _____ В.В. Волков
«26» февраля 2025 г.
Протокол УС № 2 от 26.02 2025 г.

Рабочая программа дисциплины
Машинное обучение: продвинутый уровень

образовательная программа
направление подготовки
39.04.01 Социология

направленность (профиль) программы
«Вычислительная социология»
уровень высшего образования – магистратура

Программа двух квалификаций:
- «магистр» по направлению подготовки 39.04.01 Социология;
- дополнительная квалификация – «магистр» по направлению подготовки 09.04.03 Прикладная информатика

язык обучения – русский
форма обучения - очная

Санкт-Петербург

Автор: Аркатов Д. А., к.п.н, доцент Школы вычислительных социальных наук по направлению Социология АНООВО «ЕУСПб»

Рецензент: Тенишева К.А., кандидат социологических наук, доцент направления Социология, директор программ по направлению Социология Школы Вычислительных социальных наук

Рабочая программа дисциплины **«Машинное обучение: продвинутый уровень»**, входящей в образовательную программу уровня магистратуры «Вычислительная социология», утверждена на заседании Совета Школы вычислительных социальных наук

Протокол заседания № 04 от 25.02.2025 года

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина **«Машинное обучение: продвинутый уровень»** является дисциплиной обязательной части основной профессиональной образовательной программы высшего образования «Вычислительная социология».

В рамках данного курса студенты познакомятся с основными направлениями причинно-следственного анализа в машинном обучении. В эпоху больших данных методы машинного обучения, в частности их комбинация с дата-майнингом (causality mining), начинают играть ключевую роль в ответе на сложные научные вопросы, принятии управленческих решений, объяснении поведения социальных акторов, предсказании социально-экономических событий и в других важных социальных сферах. Это станет одной из тем нашего обсуждения. Мы также покроем современные подходы к причинно-следственному выводу в машинном обучении, в частности, наработки команды В. Черножукова по двойному де-смещенному машинному обучению. Данный подход полагается на простые теоретико-статистические основания, при этом поможет слушателям применять весь спектр современных ML-методов (L1/L2, ансамбли, случайные леса и др.) для анализа данных. В конце курсах мы также покроем основные идеи, техники и решения объяснимого искусственного интеллекта (xAI) и машинного обучения для обеспечения ключевых параметров научного исследования, таких как возможность интерпретации результатов моделирования и прозрачности аналитического процесса.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы, 144 часа.

Содержание

1. НАИМЕНОВАНИЕ, ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины **«Машинное обучение: продвинутый уровень»** сформировать у обучающихся системное понимание продвинутых методов машинного обучения, включающее математическую основу моделей, принципы их построения, интерпретации и применения. Курс направлен на развитие компетенций, необходимых для проектирования, оценки и внедрения машинного обучения в исследовательскую и прикладную практику с учетом ограничений данных, производительности моделей и этических аспектов.

Задачи:

- Понять принципы построения ансамблевых моделей, методы регуляризации и особенности отбора признаков при работе с разнородными и высокоразмерными данными.
- Освоить современные методы интерпретации и объяснения предсказаний моделей, а также способы оценки доверия к выводам.
- Развить практические навыки работы с данными реального мира: категориальными, текстовыми, временными и несбалансированными выборками.
- Углубить знания по непрямому обучению, включая кластеризацию, выявление аномалий и снижение размерности с визуализацией латентных пространств.
- Научиться использовать байесовские подходы и методы автоматизированного подбора моделей и гиперпараметров с применением Python-библиотек.
- Овладеть методами оценки устойчивости и качества моделей, калибровки, построения доверительных интервалов и учета деградации моделей в продакшене.
- Сформировать представление о полном жизненном цикле ML-моделей: от подготовки данных до внедрения и мониторинга в продуктивной среде.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: общепрофессиональными (ОПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Таблица 1

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ИД.ОПК-3.1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных	Знать: принципы, методы и средства анализа и структурирования профессиональной информации З (ОПК-3)
	ИД.ОПК-3.2. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач	Уметь: интерпретировать данные и формулировать выводы и теоретические подходы для решения профессиональных задач, представляя результаты работы в виде аналитических обзоров У (ОПК-3)
	ИД.ОПК-3.3. Выявляет значимые проблемы и разрабатывает рекомендации по их решению	
	ИД.ОПК-3.4. Оформляет и представляет результаты анализа в виде аналитических обзоров	Владеть: навыками разработки рекомендаций по результатам анализа профессиональной информации В (ОПК-3)

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований	ИД.ОПК-4.1. На основе современных теорий и концепций обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний ИД.ОПК-4.2. Анализирует новые научные принципы и методы исследований в профессиональной области знаний ИД.ОПК-4.3. Применяет новые научные принципы и методы исследований в профессиональной области знаний ИД.ОПК-4.4. Разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний	Знать: актуальные направления применения новых научных принципов и методов исследований в профессиональной деятельности З (ОПК-4) Уметь: самостоятельно формировать планы и программы научных исследований с применением новых принципов и методов, характерных для выбранной отрасли науки У (ОПК-4) Владеть: навыками системного использования различных новых научных принципов и методов исследований для различных направлений науки В (ОПК-4)

В результате освоения дисциплины магистрант должен:

- знать:

- основные принципы и методы построения ансамблей моделей, регуляризации и отбора признаков;
- методы оценки качества моделей, включая калибровку вероятностей, доверительные интервалы и метрики для несбалансированных выборок;
- подходы к обнаружению аномалий, кластеризации и снижению размерности в контексте машинного обучения;
- основные этапы жизненного цикла ML-моделей: от подготовки данных до внедрения и мониторинга.

- уметь:

- настраивать и применять продвинутые алгоритмы машинного обучения к сложным и разнородным наборам данных;
- интерпретировать результаты работы моделей с помощью SHAP, LIME и других методов объяснимости;
- разрабатывать и реализовывать пайплайны обучения моделей с автоматическим подбором параметров;
- проводить моделирование и анализ временных рядов и редких событий в прикладных задачах.

- владеть:

- инструментами Python для машинного обучения, включая библиотеки sklearn, pandas, optuna, joblib, и другие;
- методами обработки и преобразования категориальных, текстовых, временных и высокоразмерных данных;
- технологиями тестирования и мониторинга моделей в продакшене с учетом drift и деградации;
- навыками критической оценки применимости моделей и их ограничений в реальных условиях.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина **«Машинное обучение: продвинутый уровень»** является дисциплиной обязательной части Блока 1 «Дисциплины (модули)» образовательной программы «Вычислительная социология». Курс читается в седьмом модуле, форма промежуточной аттестации – зачет.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения учебной и производственной практики, выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 4 зачетных единиц, 144 часа.

Таблица 2

Объем дисциплины

Типы учебных занятий и самостоятельная работа		**Объем дисциплины**									
		Всего	**Модуль**								
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Контактная работа обучающихся с преподавателем в соответствии с УП:		**28**	-	-	-	-	-	-	28	-	-	-
Лекции (Л)		**14**	-	-	-	-	-	-	14	-	-	-
Практические занятия (ПЗ)		**14**	-	-	-	-	-	-	14	-	-	-
Самостоятельная работа (СР)		**116**	-	-	-	-	-	-	116	-	-	-
Промежуточная аттестация	**форма**	**Зачет, зачет с оценкой**	-	-	-	-	-	-	Зачет	-	-	-
	час.	**-**	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		**144/4**	-	-	-	-	-	-	144/4	-	-	-

5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине: через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины

№ п/п	**Наименование тем (разделов)**	**Содержание тем (разделов)**	**Коды компетенций**	**Индикаторы компетенций** (в соот. с Таблицей 1)	**Коды ЗУВ** (в соот. с Таблицей 1)
1	Градиентный бустинг и ансамбли	Модели XGBoost, LightGBM, CatBoost; бэггинг и стеккинг; подбор гиперпараметров	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
2	Продвинутая линейная и GLM	Ridge, Lasso, ElasticNet; регуляризация; линейные модели с	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
		логарифмической и логистической функцией		ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	У (ОПК-4) В (ОПК-4)
3	Отбор и генерация признаков	Методы фильтров, обёрток и встроенных моделей; полиномиальные признаки; binning; encoding	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
4	Интерпретируемость моделей	SHAP, LIME, PDP; глобальная и локальная интерпретация; важность признаков	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
5	Работа с категориальными и текстовыми признаками	One-hot, frequency и target encoding; TF-IDF, hashing; базовая обработка текстов	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
6	Несбалансированные данные	Стратегии балансировки классов; class weights; метрики для несбалансированных задач	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
7	Кластеризация и обучение без учителя	K-means, DBSCAN, агломеративная кластеризация, silhouette score	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
8	Обнаружение аномалий	Isolation Forest, One-Class SVM, LOF; аномалии как задача новизны	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
9	Снижение размерности	PCA, ICA, t-SNE, UMAP; визуализация признаков и латентных пространств	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)
				ИД.ОПК-4.3. ИД.ОПК-4.4.	
10	Временные ряды	Feature extraction; лаги, скользящее среднее; Prophet, ML-подходы	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
11	Байесовские методы	Наивный байес; байесовские регрессии; априорные/апостериорные распределения	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
12	AutoML и подбор параметров	Auto-sklearn, optuna, hyperopt; автоматизация пайплайнов	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
13	Калибровка и доверие к моделям	Калибровка вероятностей; построение доверительных интервалов; bootstrapping	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
14	Внедрение и мониторинг	Pipeline в sklearn; joblib; FastAPI; drift и мониторинг моделей	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)

5.2 Структура дисциплины

Таблица 4

Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Л	СЗ	СРП		
Очная форма обучения							
Тема 1	Градиентный бустинг и ансамбли	10	1	1		8	Домашнее задание

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					Форма текущего контроля успеваемости*, промежуточной аттестации
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП			СР	
			Л	СЗ	СРП		
Очная форма обучения							
Тема 2	Продвинутая линейная и GLM	10	1	1		8	Домашнее задание
Тема 3	Отбор и генерация признаков	10	1	1		8	Домашнее задание
Тема 4	Интерпретируемость моделей	10	1	1		8	Домашнее задание
Тема 5	Работа с категориальными и текстовыми признаками	10	1	1		8	Домашнее задание
Тема 6	Несбалансированные данные	10	1	1		8	Домашнее задание
Тема 7	Кластеризация и обучение без учителя	10	1	1		8	Домашнее задание
Тема 8	Обнаружение аномалий	10	1	1		8	Домашнее задание
Тема 9	Снижение размерности	10	1	1		8	Домашнее задание
Тема 10	Временные ряды	10	1	1		8	Домашнее задание
Тема 11	Байесовские методы	10	1	1		8	Домашнее задание
Тема 12	AutoML и подбор параметров	10	1	1		8	Домашнее задание
Тема 13	Калибровка и доверие к моделям	12	1	1		10	Домашнее задание
Тема 14	Внедрение и мониторинг	12	1	1		10	Домашнее задание
Промежуточная аттестация 5 модуль		-	-	-	-	-	**Зачет**
Всего:		**144/4**	**14**	**14**	**-**	**116**	**-**

**Примечание: формы текущего контроля успеваемости: домашнее задание (ДЗ), практическое задание (ПрЗ)*

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных

конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1: Градиентный бустинг и ансамбли

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 2: Продвинутая линейная и GLM

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 3. Отбор и генерация признаков

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 4. Интерпретируемость моделей

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 5. Работа с категориальными и текстовыми признаками

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 6. Несбалансированные данные

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 7. Кластеризация и обучение без учителя

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 8. Обнаружение аномалий

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 9. Снижение размерности

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 10. Временные ряды

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 11. Байесовские методы

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 12. AutoML и подбор параметров

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 4 часа. Итого: 8 часов.

Тема 13. Калибровка и доверие к моделям

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 10 часов.

Тема 14. Внедрение и мониторинг

1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 4 часа.

2. Подготовка к занятиям по предложенным для обсуждения вопросам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 10 часов.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

1. Градиентный бустинг и ансамбли

- В чем принципиальное отличие бустинга от бэггинга с точки зрения bias-variance trade-off?
- Какие есть альтернативы бустингу при работе с табличными данными и когда их стоит предпочесть?
- Как работает механизм построения деревьев в LightGBM и чем он отличается от классического CART?
- Какие стратегии существуют для борьбы с переобучением в CatBoost, помимо стандартной регуляризации?

2. Продвинутая линейная и GLM

- Почему Lasso может занулять коэффициенты, а Ridge — нет?
- Как интерпретировать коэффициенты в логистической регрессии с регуляризацией?
- Чем GLM отличаются от классических линейных моделей, и какие семейства распределений можно использовать?
- Какие ограничения накладывает использование ElasticNet на коррелированные признаки?

3. Отбор и генерация признаков

- Как автоматизировать отбор признаков при помощи рекурсивных методов?
- Как влияет мультиколлинеарность на качество отбора признаков и как с ней бороться?
- В чем риски использования target encoding и как их минимизировать?
- Какие методы генерации признаков особенно эффективны для временных рядов?

4. Интерпретируемость моделей

- Как SHAP справляется с коррелированными признаками?
- Почему методы объяснимости часто дают разные оценки важности признаков?
- В чем разница между глобальной и локальной интерпретацией и когда что применять?
- Какие существуют подходы к визуализации PDP при высокой размерности признаков?

5. Работа с категориальными и текстовыми признаками

- В чем преимущества target encoding по сравнению с one-hot при большом количестве уникальных значений?
- Как можно снизить размерность TF-IDF матрицы без потери качества модели?
- Как работает hashing trick и чем он полезен в реальных задачах?
- Какие подходы к очистке текста наиболее уместны перед подачей в ML-модель?

6. Несбалансированные данные

- Почему accuracy не подходит для оценки моделей на несбалансированных выборках?
- Как реализуется метод SMOTE и какие у него ограничения?
- В каких задачах oversampling может быть хуже, чем undersampling?
- Какие альтернативы F1-мере применимы при высокой стоимости ошибок второго рода?

7. Кластеризация и обучение без учителя

- Какие существуют способы определения оптимального числа кластеров?
- Почему DBSCAN не подходит для кластеров с переменной плотностью?

- Как кластеризация может использоваться как этап предварительной обработки для задач обучения с учителем?
- Какие метрики качества кластеризации работают без знания истинных меток?

8. Обнаружение аномалий

- Как адаптировать метод One-Class SVM к большим объёмам данных?
- В чем преимущества Isolation Forest по сравнению с классическими правилами порогов?
- Можно ли использовать методы кластеризации для поиска аномалий? Как?
- Какие существуют подходы к валидации моделей аномалий без размеченных данных?

9. Снижение размерности

- Почему t-SNE плохо работает для сохранения глобальной структуры данных?
- Как сравнить качество различных методов снижения размерности?
- Какие задачи анализа данных выигрывают от использования UMAP?
- Что произойдет при применении PCA к категориальным данным?

10. Временные ряды

- Почему классические ML-модели требуют ручной генерации признаков из временных рядов?
- Какие ограничения есть у Prophet при прогнозировании резких скачков?
- Как учесть сезонность в признаках для ML-моделей?
- Чем модели с рекурсивным прогнозом отличаются от моделей с прямым прогнозом?

11. Байесовские методы

- Как изменяется интерпретация результатов при переходе от частотного подхода к байесовскому?
- В чем различие между Maximum A Posteriori и Maximum Likelihood оценками?
- Какие задачи лучше решаются байесовскими методами, чем классическими?
- Какие существуют приближения для вычисления апостериорных распределений?

12. AutoML и подбор параметров

- В чем основные отличия в стратегии работы optuna и hyperopt?
- Как формализуется пространство гиперпараметров в AutoML системах?
- Какие подходы позволяют контролировать переобучение при AutoML?
- Как AutoML учитывает необходимость предобработки данных?

13. Калибровка и доверие к моделям

- В чем разница между калиброванными и некалиброванными вероятностями предсказания?
- Какие методы калибровки лучше работают для моделей бустинга?
- Как оценить доверительный интервал предсказания модели?
- Почему плохо калиброванные модели могут быть опасны при использовании в реальных системах?

14. Внедрение и мониторинг

- Какие признаки указывают на data drift и concept drift?
- Как можно реализовать непрерывный мониторинг качества модели в бою?
- Что включает в себя тестирование модели перед внедрением?
- Какие подходы позволяют автоматизировать откат модели при ухудшении качества?

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Мэрфи, К. П. Вероятностное машинное обучение. Дополнительные темы: основания, вывод : монография / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. – Москва : ДМК Пресс, 2024. - 772 с. – ISBN 978-5-93700-120-7. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2204219
2. Шарден, Б. Крупномасштабное машинное обучение вместе с Python : практическое руководство / Б. Шарден, Л. Массарон, А. Боскетти ; пер. с анг. А. В. Логунова. - Москва : ДМК Пресс, 2018. - 360 с. - ISBN 978-5-97060-506-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2083416

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «**Машинное обучение: продвинутый уровень**» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому семинарскому занятию, активное слушание на лекциях, выполнение магистрантами домашних заданий. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по темам обсуждения.

Текущий контроль проводится в форме оценивания выполнения магистрантами письменных работ, демонстрирующих степень знакомства магистрантов с дополнительной литературой.

Таблица 5

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Градиентный бустинг и ансамбли	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Продвинутая линейная и GLM	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Отбор и генерация признаков	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Интерпретируемость моделей	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Работа с категориальными и текстовыми признаками	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Несбалансированные данные	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Кластеризация и обучение без учителя	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Обнаружение аномалий	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Снижение размерности	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Временные ряды	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Байесовские методы	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
AutoML и подбор параметров	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Калибровка и доверие к моделям	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Внедрение и мониторинг	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Домашнее задание	Магистрант выполняет работу частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные социальные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, Полное и правильное выполнение заданий работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

7.2 Контрольные задания для текущей аттестации

Примерный материал домашних заданий

1. Градиентный бустинг и ансамбли

Задание:

Используя датасет breast_cancer из sklearn.datasets, постройте модели классификации с использованием Random Forest, Gradient Boosting (из scikit-learn), XGBoost, LightGBM и CatBoost.

В ходе выполнения:

- выполните предварительную обработку данных (масштабирование, если необходимо, кодирование категориальных признаков);
- для каждой модели реализуйте подбор гиперпараметров с использованием кросс-валидации;
- сравните качество моделей по метрикам accuracy, F1 и ROC-AUC;
- постройте графики важности признаков, интерпретируйте результаты;
- сделайте выводы, какая модель наиболее устойчива и точна на этих данных.

2. Продвинутая линейная и GLM

Задание:

На основе датасета diabetes из sklearn.datasets постройте модели линейной регрессии с регуляризацией: Ridge, Lasso и ElasticNet.

Ваше задание включает:

- нормализацию всех входных признаков;
- подбор коэффициентов регуляризации с использованием GridSearchCV;
- визуализацию изменения весов признаков в зависимости от силы регуляризации;
- сравнение моделей по метрикам RMSE, MAE и R^2;
- интерпретацию того, какие признаки наиболее значимы для предсказания уровня сахара в крови.

3. Отбор и генерация признаков

Задание:

Используйте датасет AmesHousing (доступен в openml или через pycaret.datasets.get_data) для построения модели предсказания цены дома.

Необходимо:

- провести отбор признаков с помощью методов SelectKBest, RFE и Lasso;
- создать полиномиальные и взаимодействующие признаки (до второй степени);
- сравнить производительность модели градиентного бустинга на исходном и расширенном наборе признаков;
- оценить влияние отбора и генерации признаков на интерпретируемость и переобучение.

4. Интерпретируемость моделей
Задание:
На основе датасета titanic из библиотеки seaborn постройте модель классификации (LightGBM или CatBoost).
Далее:

- выполните анализ глобальной и локальной интерпретируемости модели с использованием библиотеки SHAP;
- визуализируйте зависимости типа Partial Dependence Plot (PDP) и SHAP Dependence;
- интерпретируйте влияние ключевых признаков на предсказания;
- выберите 5 объектов, для которых модель ошиблась, и проанализируйте причины этих ошибок на основе интерпретации.

5. Работа с категориальными и текстовыми признаками
Задание:
На основе датасета adult (предсказание уровня дохода, доступен через UCI) обучите модели с разными способами кодирования категориальных переменных: one-hot, target encoding, frequency encoding.
Порядок работы:

- обработайте пропущенные значения и масштабируйте числовые признаки;
- обучите LightGBM или CatBoost на каждом варианте кодировки;
- сравните производительность по ROC-AUC и log-loss;
- проанализируйте, какие признаки оказались ключевыми и как кодировка повлияла на их важность.

6. Несбалансированные данные
Задание:
Используйте датасет creditcard.csv (доступен на Kaggle, содержит реальные данные транзакций с мошенничеством).
Выполните:

- обучение базовой модели логистической регрессии;
- сравнение стратегий: class_weight='balanced', SMOTE, undersampling;
- визуализацию confusion matrix, ROC-кривой и PR-кривой;
- анализ trade-off между recall и precision, обсуждение выбора операционного порога.

7. Кластеризация и обучение без учителя
Задание:
Используйте датасет wine из sklearn.datasets, удалите метки классов и выполните кластеризацию.
Включите:

- применение алгоритмов KMeans, DBSCAN и Agglomerative Clustering;
- визуализацию кластеров после снижения размерности методом PCA;
- оценку качества кластеризации с использованием ARI и silhouette score;
- сравнение кластеров с истинными метками и интерпретацию ошибок.

8. Обнаружение аномалий
Задание:
На основе датасета forestcover (Forest CoverType, доступен через UCI ML Repository) выделите один класс как "норму", остальные — как "аномалии".
Далее:

- обучите модели One-Class SVM, Isolation Forest и Local Outlier Factor;

- сравните долю обнаруженных аномалий и визуализируйте распределение скорингов;
- выполните чувствительный анализ параметров;
- сделайте выводы о применимости методов для разных типов аномалий.

9. Снижение размерности
Задание:
Используя датасет mnist_784 (можно загрузить из openml), выполните:
- предварительное уменьшение размерности методом PCA;
- визуализацию в 2D пространстве с помощью t-SNE и UMAP;
- классификацию на основе уменьшенного представления;
- сравнение производительности модели до и после снижения размерности.

10. Временные ряды
Задание:
С использованием ежемесячных данных о пассажирообороте (AirPassengers, доступен в statsmodels.datasets) построить модель прогнозирования. Включите:
- генерацию лагов, сезонных индикаторов и тренда как признаков;
- построение модели градиентного бустинга;
- сравнение с моделью Prophet по метрикам RMSE и MAPE;
- визуализацию прогноза и остатков.

11. Байесовские методы
Задание:
Возьмите датасет sms_spam (доступен в UCI ML Repository). Реализуйте:
- наивный байесовский классификатор;
- байесовскую логистическую регрессию с использованием библиотеки PyMC;
- сравните вероятностные предсказания и оцените калиброванность моделей;
- визуализируйте апостериорные распределения коэффициентов и проанализируйте неопределённость.

12. AutoML и подбор параметров
Задание:
Используя датасет titanic, настройте модель классификации с помощью библиотеки Optuna. Реализуйте:
- определение пространства поиска гиперпараметров для XGBoost;
- настройку кросс-валидации и раннюю остановку;
- анализ влияния гиперпараметров на производительность;
- сравнение с результатами GridSearchCV.

13. Калибровка и доверие к моделям
Задание:
Используя модель классификации (например, LightGBM) на breast_cancer, выполните:
- оценку калиброванности предсказаний (без калибровки);
- применение isotonic regression и Platt scaling;
- построение unreliability диаграммы (calibration curve);
- обсуждение, как калибровка изменила поведение модели при низких и высоких вероятностях.

14. Внедрение и мониторинг

Задание:
Создайте API-сервис на основе FastAPI для модели, обученной на diabetes из sklearn. Включите:

- сохранение модели и препроцессора;
- написание эндпоинта для получения предсказаний;
- логирование входных и выходных данных;
- реализацию простой системы алертов на основе статистики входных признаков (data drift).

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет, который проходит в форме тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине.

Таблица 7

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	**Коды компетенций**	**Индикаторы компетенций** *(в соот. с Таблицей 1)*	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Критерии оценивания**	**Оценка**
зачет- тестирование	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	100-41% правильных ответов	Зачтено
				40-0% правильных ответов	Не зачтено

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а

Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценках «зачтено» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценке «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной

программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

Требования к тестам

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутвует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F=\frac{100}{K}*\left(\frac{x_1}{k_1}+\frac{x_2}{k_2}+\ldots+\frac{x_n}{k_n}\right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Примеры тестовых заданий для промежуточной аттестации

ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с выводами и рекомендациями

Задания закрытого типа (базовый уровень сложности)
Вопрос 1
Инструкция: Выберите один правильный ответ.
Вопрос: Какой метод ансамблирования дает лучшие результаты на зашумленных данных?
Варианты ответов:
a) AdaBoost
b) Gradient Boosting
c) Random Forest
d) Stacking
Поле для ответа:

Задания закрытого типа (повышенный уровень сложности)
Вопрос 1
Инструкция: Установите соответствие между элементами
Вопрос: Установите соответствие между методом машинного обучения и его ключевым свойством.

Метод	**Свойство**
1. SVM	a) Максимизирует зазор между классами
2. Random Forest	b) Автоматический отбор признаков
3. k-Means	c) Чувствителен к масштабу данных
4. Logistic Regression	d) Калибрует вероятности с sigmoid

Поле для ответа:

Задания комбинированного типа (повышенный уровень сложности)
Вопрос 1
Инструкция: Выберите **все** верные варианты ответа и приведите обоснование вашего выбора
Вопрос: Какие из следующих утверждений о методах борьбы с переобучением верны?

Варианты ответов:
a) Увеличение размера обучающей выборки всегда уменьшает переобучение
b) L1-регуляризация может приводить к отбору признаков
c) Dropout эффективен только в нейронных сетях
d) Ранняя остановка (early stopping) применима к градиентному бустингу
Поле для ответа:
Обоснование:

Вопрос 2
Инструкция: Выберите **все** верные варианты ответа и приведите обоснование.

Вопрос: Какие из перечисленных методов позволяют интерпретировать предсказания сложных моделей (например, ансамблей)?
Варианты ответов:
a) PDP (Partial Dependence Plots)
b) Значения коэффициентов линейной регрессии
c) SHAP (Shapley Additive Explanations)
d) Feature Importance в Random Forest
Поле для ответа:
Обоснование:

Задания открытого типа (высокий уровень сложности)
Вопрос 1
Инструкция: Дайте развернутый ответ на вопрос (3-4 предложения).
Вопрос: В чем преимущества использования SHAP (Shapley Additive Explanations) для интерпретации моделей машинного обучения по сравнению с традиционными методами, такими как важность признаков в Random Forest?
Развернутый ответ:

ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований

Задания закрытого типа (базовый уровень сложности)
Вопрос 1
Инструкция: Выберите один правильный ответ.\
Вопрос: Какой метод оптимизации гиперпараметров наиболее эффективен при ограниченных вычислительных ресурсах?
Варианты ответов:
a) Grid Search
b) Random Search
c) Bayesian Optimization
d) Genetic Algorithms
Поле для ответа:

Задания закрытого типа (повышенный уровень сложности)
Вопрос 1
Инструкция: Установите соответствие между элементами таблицы.
Вопрос: Установите соответствие между задачей и подходящим методом оценки качества.

Задача	**Метрика**
1. Классификация с дисбалансом	a) F1-score
2. Регрессия	b) R^2
3. Кластеризация	c) Silhouette Score
4. Ранжирование	d) NDCG

Поле для ответа:

Задания комбинированного типа (повышенный уровень сложности)

Вопрос 1

Инструкция: выберите правильные варианты ответа и приведите обоснование в отдельном поле.

Вопрос: Вам нужно решить задачу классификации изображений с ограниченным объемом размеченных данных (несколько тысяч примеров). Данные имеют высокую размерность (например, 256×256 RGB). Какие из следующих подходов являются разумными и почему?

Варианты ответов:

1. Обучение сверточной нейронной сети (CNN) с нуля на имеющихся данных.
2. Использование предобученной CNN (например, ResNet, VGG) с дообучением (fine-tuning)последних слоев.
3. Применение метода опорных векторов (SVM) с ядром RBF на сырых пикселях изображений.
4. Использование автоэнкодера для уменьшения размерности данных с последующей классификацией через Random Forest.

Обоснование:

Задания открытого типа (высокий уровень сложности)

Вопрос 1

Инструкция: Дайте развернутый ответ на вопрос (3-4 предложения).
Вопрос: Почему метод случайного леса (Random Forest) часто оказывается более устойчивым к переобучению по сравнению с отдельными деревьями решений?
Развернутый ответ:

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	**Индикаторы компетенций** *(в соот.с Таблицей 1)*	**Средства оценки** *(в соот. с Таблицами 5, 7)*
ОПК-3 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	домашнее задание, тест
ОПК-4 (ПИ)	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	домашнее задание, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки *(в соот. с Таблицами 5, 7)*	**Рекомендованный план выполнения работы**
Домашнее задание	Магистрант должен быть готовым в ходе подготовки и представления домашнего задания по темам дисциплины, выполнять следующие действия: 1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных, интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач, выявляет значимые проблемы и разрабатывает рекомендации по их решению, оформляет и представляет результаты анализа в виде аналитических обзоров 2. Обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний, анализирует новые научные принципы и методы исследований в профессиональной области знаний, применяет новые научные принципы и методы исследований в профессиональной области знаний, разрабатывает

Средства оценки *(в соот. с Таблицами 5, 7)*	Рекомендованный план выполнения работы
	предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Тест	Магистрант должен быть готовым в ходе подготовки к тесту, выполнять следующие действия: 1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных, интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач, выявляет значимые проблемы и разрабатывает рекомендации по их решению, оформляет и представляет результаты анализа в виде аналитических обзоров 2. Обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний, анализирует новые научные принципы и методы исследований в профессиональной области знаний, применяет новые научные принципы и методы исследований в профессиональной области знаний, разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.:ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2085038

2. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2023. - 990 с. - ISBN 978-5-93700-119-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/2109489

8.2. Дополнительная литература

1. Мэрфи, К. П. Вероятностное машинное обучение. Дополнительные темы: основания, вывод : монография / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. – Москва : ДМК Пресс, 2024. - 772 с. – ISBN 978-5-93700-120-7. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2204219

2. Шарден, Б. Крупномасштабное машинное обучение вместе с Python : практическое руководство / Б. Шарден, Л. Массарон, А. Боскетти ; пер. с анг. А. В. Логунова. - Москва : ДМК Пресс, 2018. - 360 с. - ISBN 978-5-97060-506-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2083416

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

1. OS Microsoft Windows (OVS OS Platform)
2. MS Office (OVS Office Platform)
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. ABBYY FineReader 11 Corporate Edition
6. ABBYY Lingvo x5
7. Adobe Acrobat Reader DC /Pro – бесплатно
8. Google Chrome – бесплатно

9. Opera – бесплатно
10. Mozilla – бесплатно
11. VLC – бесплатно
12. Яндекс Браузер

9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

Информационно-справочные системы

1. Гарант.Ру. Информационно-правовой портал: http://www.garant.ru
2. Информационная система «Единое окно доступа к образовательным ресурсам»: http://window.edu.ru/
3. Открытое образование. Ассоциация «Национальная платформа открытого образования»: http://npoed.ru
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: http://www.gov.ru
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: http://pravo.gov.ru
6. Правовой сайт КонсультантПлюс: http://www.consultant.ru/sys
7. Российское образование. Федеральный портал: http://www.edu.ru

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. Национальная электронная библиотека НЭБ: http://www.rusneb.ru
2. Президентская библиотека: http://www.prlib.ru
3. Российская государственная библиотека: http://www.rsl.ru/
4. Российская национальная библиотека: http://www.nlr.ru/poisk/

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета https://eusp.org/library/electronic-resources, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: http://elibrary.ru;
2. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: http://www.uisrussia.msu.ru/;
3. Электронные журналы по подписке (текущие номера научных зарубежных журналов).

Электронные библиотечные системы:

1. **Znanium.com** – Электронная библиотечная система **(ЭБС)** – http://znanium.com/;
2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – http://biblioclub.ru/

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета,

официальный сайт Университета (Европейский университет в Санкт-Петербурге [https://eusp.org/]),локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;
- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;
- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;
- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

ПРИЛОЖЕНИЕ 1

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

«Машинное обучение: продвинутый уровень»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому семинарскому занятию, активное слушание на лекциях, выполнение магистрантами домашних заданий. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по темам обсуждения.

Текущий контроль проводится в форме оценивания выполнения магистрантами письменных работ, демонстрирующих степень знакомства магистрантов с дополнительной литературой.

Таблица 1

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Градиентный бустинг и ансамбли	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Продвинутая линейная и GLM	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Отбор и генерация признаков	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Интерпретируемость моделей	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено

Наименование тем (разделов)	**Коды компетенций**	**Индикаторы компетенций**	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Формы текущего контроля успеваемости**	**Результаты текущего контроля**
Работа с категориальными и текстовыми признаками	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Несбалансированные данные	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Кластеризация и обучение без учителя	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Обнаружение аномалий	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Снижение размерности	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Временные ряды	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Байесовские методы	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ *(в соот. с Таблицей 1)*	Формы текущего контроля успеваемости	Результаты текущего контроля
AutoML и подбор параметров	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Калибровка и доверие к моделям	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено
Внедрение и мониторинг	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Домашнее задание	зачтено/ не зачтено

Таблица 2

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Домашнее задание	Магистрант выполняет работу частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные социальные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, Полное и правильное выполнение заданий работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

2 Контрольные задания для текущей аттестации

Материал домашних заданий,

1. Градиентный бустинг и ансамбли

Задание:

Используя датасет breast_cancer из sklearn.datasets, постройте модели классификации с использованием Random Forest, Gradient Boosting (из scikit-learn), XGBoost, LightGBM и CatBoost.

В ходе выполнения:

- выполните предварительную обработку данных (масштабирование, если необходимо, кодирование категориальных признаков);
- для каждой модели реализуйте подбор гиперпараметров с использованием кросс-валидации;
- сравните качество моделей по метрикам accuracy, F1 и ROC-AUC;

- постройте графики важности признаков, интерпретируйте результаты;
- сделайте выводы, какая модель наиболее устойчива и точна на этих данных.

2. Продвинутая линейная и GLM
Задание:
На основе датасета diabetes из sklearn.datasets постройте модели линейной регрессии с регуляризацией: Ridge, Lasso и ElasticNet.
Ваше задание включает:
- нормализацию всех входных признаков;
- подбор коэффициентов регуляризации с использованием GridSearchCV;
- визуализацию изменения весов признаков в зависимости от силы регуляризации;
- сравнение моделей по метрикам RMSE, MAE и R^2;
- интерпретацию того, какие признаки наиболее значимы для предсказания уровня сахара в крови.

3. Отбор и генерация признаков
Задание:
Используйте датасет AmesHousing (доступен в openml или через pycaret.datasets.get_data) для построения модели предсказания цены дома.
Необходимо:
- провести отбор признаков с помощью методов SelectKBest, RFE и Lasso;
- создать полиномиальные и взаимодействующие признаки (до второй степени);
- сравнить производительность модели градиентного бустинга на исходном и расширенном наборе признаков;
- оценить влияние отбора и генерации признаков на интерпретируемость и переобучение.

4. Интерпретируемость моделей
Задание:
На основе датасета titanic из библиотеки seaborn постройте модель классификации (LightGBM или CatBoost).
Далее:
- выполните анализ глобальной и локальной интерпретируемости модели с использованием библиотеки SHAP;
- визуализируйте зависимости типа Partial Dependence Plot (PDP) и SHAP Dependence;
- интерпретируйте влияние ключевых признаков на предсказания;
- выберите 5 объектов, для которых модель ошиблась, и проанализируйте причины этих ошибок на основе интерпретации.

5. Работа с категориальными и текстовыми признаками
Задание:
На основе датасета adult (предсказание уровня дохода, доступен через UCI) обучите модели с разными способами кодирования категориальных переменных: one-hot, target encoding, frequency encoding.
Порядок работы:
- обработайте пропущенные значения и масштабируйте числовые признаки;
- обучите LightGBM или CatBoost на каждом варианте кодировки;
- сравните производительность по ROC-AUC и log-loss;
- проанализируйте, какие признаки оказались ключевыми и как кодировка повлияла на их важность.

6. Несбалансированные данные

Задание:
Используйте датасет creditcard.csv (доступен на Kaggle, содержит реальные данные транзакций с мошенничеством).
Выполните:

- обучение базовой модели логистической регрессии;
- сравнение стратегий: class_weight='balanced', SMOTE, undersampling;
- визуализацию confusion matrix, ROC-кривой и PR-кривой;
- анализ trade-off между recall и precision, обсуждение выбора операционного порога.

7. Кластеризация и обучение без учителя
Задание:
Используйте датасет wine из sklearn.datasets, удалите метки классов и выполните кластеризацию.
Включите:

- применение алгоритмов KMeans, DBSCAN и Agglomerative Clustering;
- визуализацию кластеров после снижения размерности методом PCA;
- оценку качества кластеризации с использованием ARI и silhouette score;
- сравнение кластеров с истинными метками и интерпретацию ошибок.

8. Обнаружение аномалий
Задание:
На основе датасета forestcover (Forest CoverType, доступен через UCI ML Repository) выделите один класс как "норму", остальные — как "аномалии".
Далее:

- обучите модели One-Class SVM, Isolation Forest и Local Outlier Factor;
- сравните долю обнаруженных аномалий и визуализируйте распределение скорингов;
- выполните чувствительный анализ параметров;
- сделайте выводы о применимости методов для разных типов аномалий.

9. Снижение размерности
Задание:
Используя датасет mnist_784 (можно загрузить из openml), выполните:

- предварительное уменьшение размерности методом PCA;
- визуализацию в 2D пространстве с помощью t-SNE и UMAP;
- классификацию на основе уменьшенного представления;
- сравнение производительности модели до и после снижения размерности.

10. Временные ряды
Задание:
С использованием ежемесячных данных о пассажирообороте (AirPassengers, доступен в statsmodels.datasets) построить модель прогнозирования.
Включите:

- генерацию лагов, сезонных индикаторов и тренда как признаков;
- построение модели градиентного бустинга;
- сравнение с моделью Prophet по метрикам RMSE и MAPE;
- визуализацию прогноза и остатков.

11. Байесовские методы
Задание:
Возьмите датасет sms_spam (доступен в UCI ML Repository).
Реализуйте:

- наивный байесовский классификатор;
- байесовскую логистическую регрессию с использованием библиотеки PyMC;
- сравните вероятностные предсказания и оцените калиброванность моделей;
- визуализируйте апостериорные распределения коэффициентов и проанализируйте неопределённость.

12. AutoML и подбор параметров
Задание:
Используя датасет titanic, настройте модель классификации с помощью библиотеки Optuna. Реализуйте:

- определение пространства поиска гиперпараметров для XGBoost;
- настройку кросс-валидации и раннюю остановку;
- анализ влияния гиперпараметров на производительность;
- сравнение с результатами GridSearchCV.

13. Калибровка и доверие к моделям
Задание:
Используя модель классификации (например, LightGBM) на breast_cancer, выполните:

- оценку калиброванности предсказаний (без калибровки);
- применение isotonic regression и Platt scaling;
- построение unreliability диаграммы (calibration curve);
- обсуждение, как калибровка изменила поведение модели при низких и высоких вероятностях.

14. Внедрение и мониторинг
Задание:
Создайте API-сервис на основе FastAPI для модели, обученной на diabetes из sklearn. Включите:

- сохранение модели и препроцессора;
- написание эндпоинта для получения предсказаний;
- логирование входных и выходных данных;
- реализацию простой системы алертов на основе статистики входных признаков (data drift).

3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет, который проходит в форме тестирования.

Перед зачетом проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине.

Таблица 3

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	**Коды компетенций**	**Индикаторы компетенций** *(в соот. с Таблицей 1)*	**Коды ЗУВ** *(в соот. с Таблицей 1)*	**Критерии оценивания**	**Оценка**
зачет-тестирование	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4)	100-41% правильных ответов	Зачтено

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций *(в соот. с Таблицей 1)*	Коды ЗУВ *(в соот. с Таблицей 1)*	Критерии оценивания	Оценка
		ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	У (ОПК-4) В (ОПК-4)	40-0% правильных ответов	Не зачтено

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Таблица 3а

Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценках «зачтено» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе в оценке «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

4 Задания к промежуточной аттестации

Требования к тесту

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутвует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F=\frac{100}{K}*\left(\frac{x_1}{k_1}+\frac{x_2}{k_2}+\ldots+\frac{x_n}{k_n}\right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Задания к промежуточной аттестации

Тестирование

ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с выводами и рекомендациями

Задания закрытого типа (базовый уровень сложности)

Задание 1

Инструкция: Выберите один правильный ответ из предложенных.

Вопрос: Какой метод ансамблирования уменьшает дисперсию ошибки модели за счёт агрегации предсказаний множества независимо обученных моделей?

Варианты ответа:

1. Градиентный бустинг
2. Бэггинг
3. Стеккинг
4. AdaBoost

Правильный ответ:

Задание 2

Инструкция: Выберите один правильный ответ из предложенных.

Вопрос: Какой метод регуляризации линейной модели может обнулять веса незначимых

признаков?
Варианты ответа:

1. Ridge-регрессия
2. Lasso-регрессия
3. ElasticNet
4. Линейная регрессия без регуляризации

Правильный ответ:

Задание 3
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод интерпретации моделей позволяет оценить вклад каждого признака в предсказание для конкретного наблюдения?
Варианты ответа:

1. PDP (Partial Dependence Plot)
2. SHAP (SHapley Additive exPlanations)
3. Feature Importance
4. t-SNE

Правильный ответ:

Задание 4
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод кодирования категориальных признаков заменяет категорию на среднее значение целевой переменной для этой категории?
Варианты ответа:

1. One-Hot Encoding
2. Target Encoding
3. Frequency Encoding
4. Label Encoding

Правильный ответ:

Задание 5
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какая метрика НЕ подходит для оценки качества модели на несбалансированных данных?
Варианты ответа:

1. F1-score
2. ROC-AUC
3. Accuracy
4. Precision-Recall AUC

Правильный ответ:

Задание 6
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой алгоритм кластеризации использует понятие "плотности" для формирования кластеров?
Варианты ответа:

1. K-means
2. DBSCAN
3. Иерархическая кластеризация
4. GMM (Gaussian Mixture Model)

Правильный ответ:

Задание 7
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод снижения размерности сохраняет глобальную структуру данных лучше, чем t-SNE?
Варианты ответа:

1. PCA
2. UMAP
3. ICA
4. LDA

Правильный ответ:

Задание 8
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод автоматического подбора гиперпараметров использует байесовскую оптимизацию?
Варианты ответа:

1. GridSearch
2. RandomSearch
3. Optuna
4. Scikit-learn's HalvingGridSearch

Правильный ответ:

Задание 9
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод калибровки вероятностей использует изотоническую регрессию?
Варианты ответа:

1. Platt Scaling
2. Sigmoid Scaling
3. Isotonic Regression
4. Bootstrap Calibration

Правильный ответ:

Задание 10
Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой инструмент используется для мониторинга дрейфа данных (data drift) в продакшене?
Варианты ответа:

1. SHAP
2. Evidently AI
3. LIME
4. PDP

Правильный ответ:

Задания закрытого типа (повышенный уровень сложности)

Задание 11
Инструкция: Сопоставьте метод машинного обучения с его основным применением.
Вопрос:

Метод	**Применение**
1) XGBoost	A) Кластеризация данных

Метод	Применение
2) DBSCAN	B) Классификация табличных данных
3) PCA	C) Снижение размерности
4) Prophet	D) Прогнозирование временных рядов

Правильный ответ:

Задание 12
Инструкция: Сопоставьте метод интерпретации модели с его описанием.
Вопрос:

Метод	Описание
1) SHAP	A) Визуализирует усреднённое влияние признака на предсказание
2) LIME	B) Локально аппроксимирует модель линейной функцией
3) PDP	C) Оценивает вклад каждого признака на основе теории игр
4) Feature Importance	D) Ранжирует признаки по их "важности" в модели

Правильный ответ:

Задание 13
Инструкция: Сопоставьте тип регуляризации с его эффектом на веса модели.
Вопрос:

Тип регуляризации	Эффект
1) Ridge	A) Обнуляет некоторые веса
2) Lasso	B) Уменьшает веса, но не до нуля
3) ElasticNet	C) Комбинация L1 и L2 регуляризации
4) Без регуляризации	D) Не влияет на веса

Правильный ответ:

Задание 14
Инструкция: Сопоставьте метод обработки категориальных признаков с его характеристикой.
Вопрос:

Метод	Характеристика
1) One-Hot Encoding	A) Может привести к проклятию размерности
2) Target Encoding	B) Риск утечки целевой переменной
3) Frequency Encoding	C) Заменяет категорию на её частоту в данных
4) Hashing Trick	D) Фиксирует размерность признакового пространства

Правильный ответ:

Задание 15
Инструкция: Сопоставьте метрику качества с типом задачи, для которой она применяется.
Вопрос:

Метрика	Тип задачи
1) ROC-AUC	A) Регрессия
2) Silhouette Score	B) Кластеризация
3) RMSE	C) Классификация
4) F1-score	D) Несбалансированная классификация

Правильный ответ:

Задание 16
Инструкция: Сопоставьте алгоритм с его устойчивостью к выбросам.
Вопрос:

Алгоритм	Устойчивость к выбросам
1) K-means	A) Высокая
2) Isolation Forest	B) Низкая
3) DBSCAN	C) Средняя
4) Linear Regression	D) Зависит от регуляризации

Правильный ответ:

Задание 17
Инструкция: Сопоставьте метод снижения размерности с его ключевым свойством.
Вопрос:

Метод	Свойство
1) PCA	A) Сохраняет локальную структуру данных
2) t-SNE	B) Линейное преобразование
3) UMAP	C) Оптимизирован для визуализации
4) ICA	D) Разделяет сигналы на независимые компоненты

Правильный ответ:

Задание 18
Инструкция: Сопоставьте инструмент с его назначением в ML-пайплайне.
Вопрос:

Инструмент	Назначение
1) Optuna	A) Мониторинг дрейфа данных
2) Evidently	B) Подбор гиперпараметров
3) MLflow	C) Логирование экспериментов
4) FastAPI	D) Развёртывание модели как API

Правильный ответ:

Задание 19
Инструкция: Сопоставьте проблему с методом её решения.
Вопрос:

Проблема	Метод решения
1) Переобучение модели	A) SMOTE
2) Несбалансированные данные	B) Ранняя остановка (early stopping)
3) Высокая размерность данных	C) Регуляризация (L1/L2)
4) Пропущенные значения	D) PCA

Правильный ответ:

Задание 20
Инструкция: Сопоставьте этап жизненного цикла ML-модели с его задачей.
Вопрос:

Этап	Задача
1) Подготовка данных	A) Мониторинг дрейфа и деградации
2) Обучение модели	B) Построение пайплайна предобработки
3) Внедрение	C) Подбор гиперпараметров
4) Продакшн	D) Развёртывание API-сервиса

Правильный ответ:

Задания комбинированного типа (повышенный уровень сложности)

Задание 21
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: В каких случаях целесообразно использовать **LightGBM** вместо **XGBoost**?

1. При работе с небольшими данными (менее 10 тыс. строк).
2. Когда важна скорость обучения на больших данных.
3. Если требуется максимальная интерпретируемость модели.
4. При наличии категориальных признаков без предварительного кодирования.

Правильный ответ:
Обоснование:

Задание 22
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие действия помогут улучшить качество модели на **несбалансированных данных**?

1. Увеличение выборки (oversampling) миноритарного класса.
2. Использование метрики **Accuracy** для оценки.
3. Применение **class_weight='balanced'** в логистической регрессии.
4. Удаление части мажоритарного класса (undersampling).

Правильный ответ:
Обоснование:

Задание 23
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: В каких задачах **PCA** может ухудшить качество модели?

1. Когда важны нелинейные зависимости между признаками.
2. При работе с категориальными признаками без One-Hot Encoding.
3. Если исходные признаки уже имеют низкую корреляцию.
4. Когда требуется сохранить интерпретируемость признаков.

Правильный ответ:
Обоснование:

Задание 24
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие методы **не подходят** для обнаружения аномалий в данных?

1. K-means.
2. Линейная регрессия.

3. One-Class SVM.
4. Дерево решений.

Правильный ответ:
Обоснование:

Задание 25
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие утверждения о **SHAP** и **LIME** верны?

1. SHAP требует больше вычислительных ресурсов, чем LIME.
2. LIME работает только для линейных моделей.
3. SHAP обеспечивает глобальную интерпретацию модели.
4. LIME менее точен для моделей с высокой нелинейностью.

Правильный ответ:
Обоснование:

Задание 26
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: В каких случаях **ElasticNet** предпочтительнее Lasso или Ridge?

1. Когда среди признаков есть группы сильно коррелированных переменных.
2. Если требуется отбор признаков (обнуление части весов).
3. При работе с текстовыми данными после TF-IDF.
4. Когда важно сохранить все признаки без исключения.

Правильный ответ:
Обоснование:

Задание 27
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **гиперпараметры CatBoost** критично настраивать для борьбы с переобучением?

1. learning_rate.
2. depth.
3. l2_leaf_reg.
4. random_seed.

Правильный ответ:
Обоснование:

Задание 28
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **метрики** стоит использовать для оценки модели прогнозирования **временных рядов**?

1. RMSE.
2. Accuracy.
3. MAPE.
4. ROC-AUC.

Правильный ответ:
Обоснование:

Задание 29
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **подходы** помогут улучшить качество модели на **текстовых данных**?

1. Уменьшение размерности TF-IDF через SVD.
2. Использование Stop Words.

3. Применение Word2Vec вместо Bag of Words.
4. One-Hot Encoding для каждого слова.

Правильный ответ:
Обоснование:

Задание 30
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **этапы** обязательны при внедрении ML-модели в **продакшн**?

1. Логирование входных данных и предсказаний.
2. Калибровка вероятностей.
3. Настройка мониторинга дрейфа данных.
4. Визуализация SHAP-значений для каждого запроса.

Правильный ответ:
Обоснование:

Задание 31
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **алгоритмы** могут использоваться для **понижения размерности** без потери интерпретируемости?

1. PCA.
2. t-SNE.
3. UMAP.
4. Факторный анализ.

Правильный ответ:
Обоснование:

Задание 32
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **стратегии** эффективны для борьбы с **переобучением** в нейронных сетях?

1. Добавление Dropout-слоёв.
2. Увеличение количества эпох обучения.
3. Использование L2-регуляризации.
4. Уменьшение размера батча.

Правильный ответ:
Обоснование:

Задание 33
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **утверждения** о **Prophet** верны?

1. Подходит только для стационарных временных рядов.
2. Автоматически учитывает сезонность и праздники.
3. Требует ручной настройки лагов.
4. Хорошо работает с пропусками в данных.

Правильный ответ:
Обоснование:

Задание 34
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **методы** позволяют оценить **неопределённость предсказаний** байесовской модели?

1. Доверительные интервалы для параметров.
2. Bootstrap-агрегирование (Bagging).

3. Построение апостериорного распределения.
4. Кросс-валидация.

Правильный ответ:
Обоснование:

Задание 35
Инструкция: Выберите **один или несколько** правильных ответов.
Вопрос: Какие **подходы** помогают выявить **data drift** в продакшене?

1. Сравнение распределений признаков между обучающей и текущей выборками.
2. Мониторинг метрик качества модели (например, Accuracy).
3. Визуализация SHAP-значений для новых данных.
4. Анализ изменения важности признаков.

Правильный ответ:
Обоснование:

Задания открытого типа (высокий уровень сложности)

Задание 36
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие критерии следует учитывать при выборе между градиентным бустингом (например, XGBoost) и нейронной сетью для задачи классификации табличных данных? Обоснуйте свой ответ.
Поле для ответа:

Задание 37
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Почему при работе с временными рядами классические ML-модели (например, Random Forest) требуют ручной генерации признаков, в отличие от специализированных методов (например, ARIMA или Prophet)?
Поле для ответа:

Задание 38
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие шаги необходимо выполнить для корректного внедрения ML-модели в продакшн, чтобы минимизировать риски её деградации?
Поле для ответа:

Задание 39
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: В чем преимущества и недостатки использования AutoML (например, Auto-Sklearn или H2O) по сравнению с ручным подбором моделей и гиперпараметров?
Поле для ответа:

Задание 40
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно улучшить качество модели, если её предсказания плохо калиброваны (например, вероятности занижены или завышены)?
Поле для ответа:

Задание 41

Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы интерпретации модели следует выбрать, если важно объяснить предсказание для конкретного наблюдения, а не глобальное поведение модели?
Поле для ответа:

Задание 42
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Почему при обработке текстовых данных методом TF-IDF часто применяют уменьшение размерности (например, SVD)? Какие альтернативы можно использовать?
Поле для ответа:

Задание 43
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие стратегии балансировки классов в несбалансированных данных могут ухудшить качество модели и почему?
Поле для ответа:

Задание 44
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно использовать кластеризацию в качестве этапа предобработки данных для задачи обучения с учителем?
Поле для ответа:

Задание 45
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Почему в задачах обнаружения аномалий метрики типа Accuracy неэффективны и какие альтернативы следует использовать?
Поле для ответа:

Задание 46
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие особенности данных временных рядов осложняют применение кросс-валидации в классическом виде?
Поле для ответа:

Задание 47
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие практические шаги следует предпринять, если модель в продакшене начала резко терять качество из-за concept drift?
Поле для ответа:

Задание 48
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: В чем преимущества байесовских методов перед частотным подходом в машинном обучении?
Поле для ответа:

Задание 49
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы позволяют снизить риск переобучения при работе с градиентным бустингом (например, XGBoost)?
Поле для ответа:

Задание 50
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Почему при работе с NLP предобученные языковые модели (например, BERT) часто эффективнее, чем традиционные методы (TF-IDF, Word2Vec)?
Поле для ответа:

ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований

Задания закрытого типа (базовый уровень)

Задание 51

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод позволяет эффективно обрабатывать категориальные признаки с большим количеством уникальных значений без значительного увеличения размерности данных?
Варианты ответа:

1. One-Hot Encoding
2. Target Encoding
3. Label Encoding
4. Frequency Encoding
   Правильный ответ:

Задание 52

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой алгоритм машинного обучения наиболее подходит для обработки текстовых данных с сохранением семантических связей между словами?
Варианты ответа:

1. TF-IDF + Logistic Regression
2. Word2Vec + Random Forest
3. BERT
4. CountVectorizer + Naive Bayes
   Правильный ответ:

Задание 53

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод позволяет эффективно комбинировать преимущества L1 и L2 регуляризации в линейных моделях?
Варианты ответа:

1. Ridge Regression
2. Lasso Regression
3. ElasticNet
4. Polynomial Regression
   Правильный ответ:

Задание 54

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой подход наиболее эффективен для обработки временных рядов с пропущенными значениями?
Варианты ответа:

1. Замена пропусков нулями
2. Линейная интерполяция
3. Удаление строк с пропусками
4. Замена средним значением по всему ряду
   Правильный ответ:

Задание 55

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод позволяет эффективно снижать размерность данных при сохранении нелинейных зависимостей между признаками?
Варианты ответа:

1. PCA
2. t-SNE
3. LDA
4. Factor Analysis
   Правильный ответ:

Задание 56

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой алгоритм наиболее подходит для обнаружения аномалий в многомерных данных без предварительного обучения на размеченных примерах?
Варианты ответа:

1. K-Nearest Neighbors
2. Isolation Forest
3. Support Vector Machines
4. Decision Trees
   Правильный ответ:

Задание 57

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод позволяет эффективно комбинировать предсказания нескольких моделей для улучшения итогового результата?
Варианты ответа:

1. Bagging
2. Boosting
3. Stacking
4. Dropout
   Правильный ответ:

Задание 58

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой подход позволяет автоматически определять оптимальное количество кластеров в алгоритме K-means?
Варианты ответа:

1. Метод локтя (Elbow Method)
2. Случайный выбор
3. Фиксированное значение k=5
4. Линейная регрессия
   Правильный ответ:

Задание 59

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой метод позволяет эффективно обрабатывать несбалансированные данные без изменения распределения исходных классов?
Варианты ответа:

1. SMOTE
2. Undersampling
3. Class Weighting
4. Удаление миноритарного класса
   Правильный ответ:

Задание 60

Инструкция: Выберите один правильный ответ из предложенных.
Вопрос: Какой инструмент позволяет эффективно развертывать ML-модели в виде REST API?
Варианты ответа:

1. Jupyter Notebook
2. FastAPI
3. Pandas
4. Matplotlib
   Правильный ответ:

Задания закрытого типа (повышенный уровень сложности)

Задание 61
Инструкция: Сопоставьте метод обработки пропущенных значений с его описанием.
Вопрос:

Метод обработки	**Описание**
1) Медианная замена	A) Удаление строк с пропусками
2) KNN Imputer	B) Замена на наиболее часто встречающееся значение

Метод обработки	Описание
3) Полное удаление	C) Замена на значение ближайших соседей
4) Модовая замена	D) Замена на центральное значение распределения

Правильный ответ:

Задание 62
Инструкция: Сопоставьте тип нейронной сети с областью её применения.
Вопрос:

Тип нейронной сети	Применение
1) CNN (Сверточная)	A) Обработка временных рядов
2) RNN (Рекуррентная)	B) Обработка изображений
3) GAN (Генеративная)	C) Генерация новых данных
4) Transformer	D) Обработка естественного языка

Правильный ответ:

Задание 63
Инструкция: Сопоставьте метод оптимизации с его характеристикой.
Вопрос:

Метод оптимизации	Характеристика
1) SGD	A) Адаптивный шаг обучения
2) Adam	B) Импульс для ускорения сходимости
3) RMSprop	C) Простейший градиентный спуск
4) Momentum	D) Адаптация скорости для каждого параметра

Правильный ответ:

Задание 64
Инструкция: Сопоставьте метрику качества с типом задачи машинного обучения.
Вопрос:

Метрика	Тип задачи
1) Mean Absolute Error	A) Классификация текста
2) BLEU Score	B) Регрессия

Метрика	Тип задачи
3) Intersection over Union	C) Машинный перевод
4) Perplexity	D) Обнаружение объектов

Правильный ответ:

Задание 65
Инструкция: Сопоставьте библиотеку Python с её основным назначением.
Вопрос:

Библиотека	Назначение
1) NLTK	A) Компьютерное зрение
2) OpenCV	B) Обработка естественного языка
3) Dask	C) Параллельные вычисления
4) Plotly	D) Интерактивная визуализация

Правильный ответ:

Задание 66
Инструкция: Сопоставьте метод аугментации данных с типом данных.
Вопрос:

Метод аугментации	Тип данных
1) Random Cropping	A) Текстовые данные
2) Synonym Replacement	B) Временные ряды
3) Time Warping	C) Изображения
4) Gaussian Noise	D) Табличные данные

Правильный ответ:

Задание 67
Инструкция: Сопоставьте алгоритм с его вычислительной сложностью.
Вопрос:

Алгоритм	Сложность
1) K-means	A) $O(n^2)$
2) Decision Tree	B) $O(n \log n)$

Алгоритм	Сложность
3) DBSCAN	C) $O(n^3)$
4) SVM	D) O(n)

Правильный ответ:

Задание 68
Инструкция: Сопоставьте метод регуляризации с его эффектом на модель.
Вопрос:

Метод регуляризации	Эффект
1) Dropout	A) Уменьшение весов
2) Batch Normalization	B) Случайное отключение нейронов
3) L1 Regularization	C) Нормализация активаций
4) Early Stopping	D) Прекращение обучения при переобучении

Правильный ответ:

Задание 69
Инструкция: Сопоставьте метод объяснения моделей с его особенностью.
Вопрос:

Метод объяснения	Особенность
1) LIME	A) Глобальная интерпретация
2) SHAP	B) Локальная аппроксимация
3) PDP	C) Теория игр
4) Feature Importance	D) Усредненное влияние признаков

Правильный ответ:

Задание 70
Инструкция: Сопоставьте технологию с её применением в MLOps.
Вопрос:

Технология	Применение
1) Kubeflow	A) Мониторинг моделей
2) MLflow	B) Оркестрация пайплайнов

Технология **Применение**

3) Prometheus C) Логирование экспериментов

4) Airflow D) Планирование задач

Правильный ответ:

Задания комбинированного типа (повышенный уровень сложности)

Задание 71
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие подходы к обработке категориальных признаков являются предпочтительными при использовании градиентного бустинга на больших табличных датасетах с высокой кардинальностью признаков?
Варианты ответа:
A. One-hot encoding
B. Frequency encoding
C. Target encoding
D. Label encoding
Правильный ответ:
Обоснование:

Задание 72
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие методы наиболее эффективно применимы для интерпретации сложных моделей в задаче медицинской диагностики?
Варианты ответа:
A. SHAP
B. Partial Dependence Plot (PDP)
C. Cross-validation
D. RandomizedSearchCV
Правильный ответ:
Обоснование:

Задание 73
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие действия необходимо предпринять перед применением PCA для анализа данных?
Варианты ответа:
A. Масштабировать числовые признаки
B. Удалить все категориальные признаки
C. Заменить пропущенные значения средними
D. Провести отбор признаков по важности
Правильный ответ:
Обоснование:

Задание 74
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие методы наиболее целесообразно использовать при построении модели обнаружения аномалий в потоковых данных?
Варианты ответа:
A. Isolation Forest
B. DBSCAN
C. One-Class SVM
D. Rolling statistics с Z-оценкой
Правильный ответ:
Обоснование:

Задание 75
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
В каком случае стоит отдать предпочтение метрике PR-AUC вместо ROC-AUC?
Варианты ответа:
A. Когда классы сбалансированы
B. При наличии большого количества негативных примеров
C. В задачах с высоким дисбалансом классов
D. Если модель выдает вероятности, а не метки
Правильный ответ:
Обоснование:

Задание 76
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Каковы преимущества использования Optuna для подбора гиперпараметров?
Варианты ответа:
A. Поддержка байесовской оптимизации
B. Возможность использовать параллельные вычисления
C. Не требует задания целевой метрики
D. Позволяет автоматизировать раннюю остановку
Правильный ответ:
Обоснование:

Задание 77
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие признаки наиболее полезны при построении модели прогнозирования на временных рядах?
Варианты ответа:
A. Лаговые значения
B. Категориальные признаки
C. Скользящее среднее
D. One-hot кодирование даты
Правильный ответ:
Обоснование:

Задание 78
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие методы являются подходящими для оценки доверительных интервалов предсказаний модели?
Варианты ответа:
A. Bootstrap
B. Dropout в режиме теста
C. SHAP values
D. Байесовский вывод
Правильный ответ:
Обоснование:

Задание 79
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
В каких случаях использование ансамблей моделей может оказаться неэффективным?
Варианты ответа:
A. Малый объем обучающей выборки
B. Высокая корреляция между моделями в ансамбле
C. Использование моделей с низкой вариативностью
D. Когда важна интерпретируемость
Правильный ответ:
Обоснование:

Задание 80
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие особенности характерны для применения методов SHAP?
Варианты ответа:
A. SHAP применим только к линейным моделям
B. Он учитывает взаимодействия между признаками
C. Он может быть использован для локальной интерпретации
D. Требует доступа к внутренней структуре модели
Правильный ответ:
Обоснование:

Задание 81
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Что может быть причиной data drift в продуктивной ML-системе?
Варианты ответа:
A. Изменение распределения входных данных
B. Снижение точности модели
C. Появление новых категорий в признаках
D. Увеличение объема данных для обучения
Правильный ответ:
Обоснование:

Задание 82
Инструкция: Выберите один или несколько правильных ответов из предложенных.

Вопрос:
Какие методы кластеризации подвержены влиянию формы и плотности кластеров?
Варианты ответа:
A. KMeans
B. DBSCAN
C. Agglomerative Clustering
D. Gaussian Mixture Model
Правильный ответ:
Обоснование:

Задание 83
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Как можно минимизировать переобучение при использовании AutoML?
Варианты ответа:
A. Использовать ограничение на количество моделей
B. Применять кросс-валидацию
C. Ограничить глубину дерева
D. Повысить количество признаков
Правильный ответ:
Обоснование:

Задание 84
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Что позволяет визуализация латентного пространства после применения UMAP?
Варианты ответа:
A. Выявить кластеры в данных
B. Улучшить классификацию модели
C. Определить выбросы
D. Интерпретировать вклад каждого признака
Правильный ответ:
Обоснование:

Задание 85
Инструкция: Выберите один или несколько правильных ответов из предложенных.
Вопрос:
Какие особенности имеет внедрение ML-моделей через FastAPI?
Варианты ответа:
A. Возможность логирования входов и предсказаний
B. Высокая скорость развертывания
C. Автоматический подбор гиперпараметров
D. Легкость интеграции с мониторингом
Правильный ответ:
Обоснование:

Задания открытого типа (высокий уровень сложности)

Задание 86

Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие стратегии можно использовать для эффективного обучения модели на небольших наборах данных? Обоснуйте выбор каждой стратегии.

Поле для ответа:

Задание 87

Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно улучшить производительность модели при работе с высокоразмерными данными (тысячи признаков)?

Поле для ответа:

Задание 88

Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы позволяют обнаруживать и устранять смещение (bias) в тренировочных данных?

Поле для ответа:

Задание 89

Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие подходы позволяют эффективно комбинировать несколько моделей машинного обучения?

Поле для ответа:

Задание 90
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно адаптировать модель машинного обучения для работы в условиях концептуального дрейфа (concept drift)?
Поле для ответа:

Задание 91
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы позволяют оценивать неопределенность предсказаний в байесовских моделях?
Поле для ответа:

Задание 92
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно улучшить интерпретируемость сложных моделей (нейросети, бустинг) без существенной потери точности?
Поле для ответа:

Задание 93
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие стратегии эффективны для обработки мультимодальных данных (текст + изображения + табличные данные)?

Поле для ответа:

Задание 94
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы позволяют эффективно сравнивать производительность разных моделей машинного обучения?
Поле для ответа:

Задание 95
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно оптимизировать процесс подбора гиперпараметров для сложных моделей?
Поле для ответа:

Задание 96
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие подходы позволяют эффективно развертывать большие ML-модели в production?
Поле для ответа:

Задание 97
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие методы позволяют обрабатывать временные ряды с переменной частотой дискретизации?
Поле для ответа:

Задание 98
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно улучшить устойчивость модели к атакам adversarial examples?
Поле для ответа:

Задание 99
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Какие подходы позволяют эффективно обучать модели на данных с частичной разметкой?
Поле для ответа:

Задание 100
Инструкция: Дайте развернутый ответ на поставленный вопрос.
Вопрос: Как можно оценить экономический эффект от внедрения ML-модели в бизнес-процессы?
Поле для ответа:

5 Средства оценки индикаторов достижения компетенций

Таблица 4

Средства оценки индикаторов достижения компетенций

Коды компетенций	**Индикаторы компетенций** *(в соот.с Таблицей 1)*	**Средства оценки** *(в соот. с Таблицами 5, 7)*
ОПК-3 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3.	домашнее задание, тест

	ИД.ОПК-3.4.	
ОПК-4 (ПИ)	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	домашнее задание, тест

Таблица 5

Описание средств оценки индикаторов достижения компетенций

Средства оценки *(в соот. с Таблицами 5, 7)*	**Рекомендованный план выполнения работы**
Домашнее задание	Магистрант должен быть готовым в ходе подготовки и представления домашнего задания по темам дисциплины, выполнять следующие действия: 1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных, интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач, выявляет значимые проблемы и разрабатывает рекомендации по их решению, оформляет и представляет результаты анализа в виде аналитических обзоров 2. Обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний, анализирует новые научные принципы и методы исследований в профессиональной области знаний, применяет новые научные принципы и методы исследований в профессиональной области знаний, разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Тест	Магистрант должен быть готовым в ходе подготовки к тесту, выполнять следующие действия: 1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных, интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач, выявляет значимые проблемы и разрабатывает рекомендации по их решению, оформляет и представляет результаты анализа в виде аналитических обзоров 2. Обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний, анализирует новые научные принципы и методы исследований в профессиональной области знаний, применяет новые научные принципы и методы исследований в профессиональной области знаний, разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний