

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
Московский государственный университет имени М.В. Ломоносова  
Факультет вычислительной математики и кибернетики

**УТВЕРЖДАЮ**  
декан факультета вычислительной  
математики и кибернетики  
**И.А. Соколов /**  
«27» сентября 2023г.



**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**  
по дисциплине  
**Интерпретируемое машинное обучение**

---

**Уровень высшего образования:**  
бакалавриат

**Направление подготовки / специальность:**  
02.03.02 "Фундаментальная информатика и информационные технологии" (3++)

**Направленность (профиль) ОПОП:**  
Искусственный интеллект и анализ данных

**Форма обучения:**  
очная

Рассмотрен и утвержден  
на заседании Ученого совета факультета ВМК  
(протокол №7, от 27 сентября 2023 года)

Москва 2023

## 1. ФОРМЫ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В процессе и по завершении изучения дисциплины оценивается формирование у студентов следующих компетенций:

<b>Планируемые результаты обучения по дисциплине (модулю)</b>		
<b>Содержание и код компетенции.</b>	<b>Индикатор (показатель) достижения компетенции</b>	<b>Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций</b>
ПК-5. Способен использовать инструментальные средства для решения задач машинного обучения	<p>ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи</p> <p>ПК-5.2. Разрабатывает модели машинного обучения для решения задач</p> <p>ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения</p>	<p>ПК-5.1. 3-1. Знает возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения</p> <p>ПК-5.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения</p> <p>ПК-5.2.3-1. Знает функциональные возможности современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения</p> <p>ПК-5.2. 3-2. Знает принципы проведения машинного эксперимента, проблемы переобучения и недообучения модели, требования к обучающей, тестовой и валидационной выборкам для решения задач анализа данных и машинного обучения</p> <p>ПК-5.2. У-1. Умеет применять современные инструментальные средства и системы программирования для разработки моделей машинного обучения</p> <p>ПК-5.2. У-2. Умеет планировать и выполнять машинные эксперименты, оценивать точность и качество построенных моделей</p> <p>ПК-5.3. 3-1. Знает принципы</p>

		<p>построения систем искусственного интеллекта, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта с применением машинного обучения</p> <p>ПК-5.3. 3-2. Знает методологию проведения массово параллельных вычислений для ускорения машинного обучения (с использованием GPU)</p> <p>ПК-5.3. 3-3. Знает принципы работы распределенных кластерных систем</p> <p>ПК-5.3. У-1. Умеет решать задачи по выполнению коллективной проектной деятельности для создания, поддержки и использования системы искусственного интеллекта с применением машинного обучения и массово параллельных вычислений для ускорения машинного обучения</p> <p>ПК-5.3. У-2. Умеет работать с распределенной кластерной системой при создании, поддержке и использовании систем искусственного интеллекта</p>
--	--	--

### 1.1. Текущий контроль успеваемости

Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

Примеры заданий для практических занятий

**Задача 1.** Интерпретация для задачи регрессии. Задан датасет, содержащий в себе информацию о весе и росте 25000 человек. Обучить модель, построить уравнение линейной регрессии вида  $weight = \beta_0 + \beta_1 height$ . Визуализировать обучающую выборку вместе с полученным уравнением. Посчитать MAE, MSE, коэффициент корреляции Пирсона. Объяснить полученные значения.

**Задача 2.** Поиск ассоциативных правил с помощью ПП SPMF:

(а) Для массива данных о контекстной рекламе 2000 компаний  $\times$  3000 словосочетаний найти ассоциативные правила для минимальной поддержки  $minsupp = 35$  и  $minconf = 1$ . Необходимо указать число таких правил.

(b) Для исходного массива данных найти замкнутые ассоциативные правила для минимальной поддержки  $\text{minsupp}=35$  и  $\text{minconf}=1$ . Необходимо указать число таких правил.

(c) Для исходного массива данных найти 5 самых частых правил при минимальной достоверности  $\text{minconf} = 0,8$ . Необходимо указать эти правила и дать интерпретацию.

**Задача 3.** Анализ посещаемости сайтов на основе решеток формальных понятий:

Для трех контекстов о посещаемости некоторого сайта в терминах посещений сайтов новостной, образовательной и финансовой тематики необходимо выполнить:

(a) Удалением некоторого числа сайтов (признаков) или пользователей (объектов) добиться числа формальных понятий не менее 100, но не сильно превышающего это значение.

(b) Для контекстов, полученных удалением объектов или признаков в пункте а), построить диаграммы решеток понятий.

(c) Привести 3–5 примеров понятий в виде пары (размер объема понятия, содержание понятия) для размера содержания 2 и более сайта. Дать содержательную интерпретацию найденных понятий.

(d) Привести пример импликации вида  $A \rightarrow B$ , найденной по диаграмме решетки понятий с указанием ее поддержки.

**Задача 4.** Интерпретация для задачи классификации. Задан датасет о сердечно-сосудистых заболеваниях, 11 признаков, 70000 объектов. Обучить модель логистической регрессии.

Определить, какие из признаков наиболее пагубно влияют на развитие сердечно-сосудистых заболеваний (глобальная интерпретация). Для конкретного человека определить, склонен ли он к сердечно-сосудистым заболеваниям (локальная интерпретация). Определить значения для каждого из признаков, для которых вероятность иметь сердечно-сосудистые заболевания превышает 0.5. Визуализировать полученные граничные значения вместе с обучающей выборкой.

**Задача 5.** Сравнение метрик качества для задач регрессии. Задан датасет данных об опозданиях самолетов некоторой авиакомпании. Построить модель, позволяющую узнать, на сколько минут опоздает самолет. Обучить регрессоры:

- линейная регрессия
- полиномиальная регрессия
- полиномиальная регрессия без квадратов
- Ridge регрессия
- дерево решений
- регуляризованная линейная регрессия kNN
- случайный лес
- MLP регрессия

Сравнить полученные модели по метрикам RMSE,  $R^2$ -score.

**Задача 6.** Сравнение метрик качества для задач классификации. Задан датасет данных об опозданиях самолетов некоторой авиакомпании. Построить модель, позволяющую узнать, опоздает ли самолет. Обучить классификаторы:

- логистическая регрессия
- RidgeClassifierCV
- DecisionTreeClassifier
- KNeighborsClassifier
- GaussianNB
- GradientBoostingClassifier
- RandomForestClassifier
- MLPClassifier

Сравнить полученные модели по метрикам: accuracy, recall, ROC-AUC, F1-score, коэффициент корреляции Мэттьюса.

**Задача 7.** Сравнение метрик качества для методов сокращения размерности. Задан некоторый датасет. Необходимо сократить его размерность до 3, используя методы: PCA, t-SNE, VAE.

Визуализировать и сравнить полученные результаты.

**Задача 8.** Изучение feature importance. Задан датасет, описывающий свойства характера людей в зависимости от того, каким ребенком был испытуемый (старший, младший, средний), определить на какие из свойств характера наиболее влияет очередность рождения. Использовать классификаторы:

- DecisionTreeClassifier
- GradientBoostingClassifier
- RandomForestClassifier
- LogisticRegression
- LinearDiscriminantAnalysis
- MLPClassifier

Для каждой из моделей определить feature importance. Вычислить PFI. Визуализировать графики PDP и ICE, провести их сравнительный анализ с методами поиска feature importance для известной модели.

## 1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме экзамена

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

## 1.3. Типовые задания для проведения промежуточной аттестации

Список вопросов

1. Ключевые аспекты интерпретируемости. Открытые задачи в области интерпретируемого машинного обучения.
2. Оценка качества интерпретации модели обучения и способы ее получения.
3. Интерпретируемые модели. Линейная регрессия. Логистическая регрессия.
4. Интерпретируемые модели. GLM, GAM. Деревья решений.
5. Интерпретируемые модели. ассоциативные правила. Алгоритм RuleFit.
6. Интерпретируемые модели. Наивный байесовский классификатор. Knn.
7. Методы интерпретации независимые от модели. PDP. ICE. ALE.
8. Методы интерпретации независимые от модели. SHAP. Глобальные методы.
9. Методы интерпретации независимые от модели. Локальные методы. LIME.
10. Интерпретация с помощью примеров.
11. Правдоподобные и противоречащие объяснения.
12. Методы интерпретации для многомерного прогнозирования и анализа чувствительности.
13. Feature selection и feature engineering для задачи интерпретации.
14. Методы устранения предвзятости. Методы определения причинно-следственных связей.
15. Способы модификации модели для лучшей интерпретируемости.
16. Обеспечение надежности. Состязательная робастность.

## 2. КРИТЕРИИ ОЦЕНКИ ПО ДИСЦИПЛИНЕ

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
виды оценочных средств				
<b>Знания</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
<b>Умения</b> (виды оценочных средств: приведены в п. 1.2.)	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности не принципиального характера)	Успешное и систематическое умение
<b>Навыки</b> (владения, опыт деятельности) (виды оценочных средств: приведены в п. 1.2..)	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач