

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

Анализ и визуализация данных

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:

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-7. Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	<p>ПК-7.1. Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>ПК-7.2. Выполняет подготовку и разметку структурированных и неструктурированных данных для машинного обучения</p>	<p>ПК-7.1.3-1. Знает виды представления данных, методы поиска и парсинга данных.</p> <p>ПК-7.1. 3-2. Знает уровни представления данных (ODS, DDL, семантический слой, модель данных)</p> <p>ПК-7.1. 3-3. Знает основные инструменты, библиотеки и технологии Data Science</p> <p>ПК-7.1. У-1. Умеет отделять достоверные источники данных от сомнительных, осуществлять критических отбор данных, проверять их на целостность и непротиворечивость</p> <p>ПК-7.1. У-2. Умеет использовать инструменты и библиотеки для Data Science для поиска данных в открытых источниках, специализированных библиотеках и репозиториях</p> <p>ПК-7.2. 3-1. Знает методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных</p> <p>ПК-7.2. 3-2. Знает методы планирования вычислительного эксперимента, формирования обучающей и контрольной выборок</p> <p>ПК-7.2. У-1. Умеет выявлять и исключать из массива данных ошибочные данные и выбросы</p>

		<p>ПК-7.2. У-2. Умеет выделять входные и выходные переменные с целью использования предиктивных моделей</p> <p>ПК-7.2. У-3. Умеет осуществлять разметку структурированных и неструктурированных данных</p> <p>ПК-7.2. У-4. Умеет использовать инструменты, библиотеки и технологии Data Science для подготовки и разметки структурированных и неструктурированных данных для машинного обучения</p> <p>ПК-7.2. У-5. Умеет использовать методы и технологии массово параллельной обработки и анализа данных</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:

решение индивидуальных заданий

Пример индивидуального задания

1. Содержательная задача

Набор данных содержит информацию о клиентах банка и предыдущей маркетинговой кампании банка.

Банк хочет повысить свой депозитный портфель путем предложения предыдущим клиентам срочного вклада.

На входе мы имеем профиль клиента, которому уже предлагался вклад и каков был результат предложения: сделал или не сделал вклад.

Цель заключается в том, чтобы предсказать, откликнется ли клиент на предложение банка по срочному вкладу или нет.

Желательно минимизировать количество предложений вклада, которые заведомо не будут приняты клиентом так как это повышает расходы на маркетинг кроме того большое количество неэффективных предложений может привести к репутационным потерям банка.

2. Данные

Этот набор данных взят из общедоступного источника.

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, In press, <http://dx.doi.org/10.1016/j.dss.2014.03.001>

Авторы: Sérgio Moro (ISCTE-IUL), Paulo Cortez (Univ. Minho) and Paulo Rita (ISCTE-IUL) @ 2014

Варианты данных: Банковский маркетинг (bank-full.csv)

Подробности и описание <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

3. Формальная задача

Дано: данные маркетингового исследования (профиль клиента) и результат взаимодействия клиента с банком.

Цель бинарной классификации - предсказать, подпишет ли клиент банковский депозит (переменная y).

Показатель качества: минимизация неэффективных предложений банковского продукта.

Необходимо определить какой из классификаторов (LinearDiscriminantAnalysis, QuadraticDiscriminantAnalysis, KNeighborsClassifier, LogisticRegression, GaussianNB) дает наилучшую модель машинного обучения.

Содержание отчета (в формате PDF):

ФИО полностью, факультет, курс, группа

Постановка задачи

Описание классификаторов (кратко)

Текст программы

Пример работы программы в виде скриншота + графики и диаграммы

Выводы по результатам выбора классификатора

1.2. Промежуточная аттестация

Промежуточная аттестация осуществляется в форме зачет

В качестве средств, используемых на промежуточной аттестации предусматривается:

Билеты

1.3. Типовые задания для проведения промежуточной аттестации

Вопросы к зачету

1. Классификация задач интеллектуального анализа данных.
2. Расстояния, сходства, метрики для различных типов данных.
3. Этапы подготовки данных. Выдвижение гипотез.
4. Методы сбора и систематизации фактов.
5. Методы проведения экспертиз для выявления наиболее значимых факторов.
6. Понятия парциальной и комплексной обработки.
7. Анализ качества полученных моделей.
8. Введение в NumPy. Векторизованные вычисления. Агрегирующие функции. Broadcasting. Логические операции.
9. Введение в pandas. Индексы в pandas. Иерархические (множественные) индексы. Иерархические колонки.
10. Агрегирование в pandas. Объединение нескольких наборов данных. Решение базовых задач анализа данных.
11. Эвристические алгоритмы кластерного анализа. Иерархическая кластеризация. Числовые характеристики кластерного разбиения.
12. Байесовский подход к задаче классификации. Многомерное гауссовское распределение. Линейный и квадратичный дискриминантный анализ.
13. Однофакторная линейная регрессионный.
14. Множественная линейная регрессия.
15. Проверка гипотез о коэффициентах регрессии.
16. Нелинейная регрессия. Логистическая регрессия.

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				
Знания (виды оценочных средств: приведены в п. 1.2.)	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения (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
Навыки (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