

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
Обнинский институт атомной энергетики –
филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(ИАТЭ НИЯУ МИФИ)

ОТДЕЛЕНИЕ ЯДЕРНОЙ ФИЗИКИ И ТЕХНОЛОГИЙ

Одобрено на заседании
Ученого совета ИАТЭ НИЯУ
МИФИ
Протокол от 24.04.2023 № 23.4

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ по дисциплине

Машинное обучение и большие данные

название дисциплины

для направления подготовки

12.03.01 Приборостроение

код и название направления подготовки

образовательная программа

Приборы и методы контроля качества и диагностики

Форма обучения: заочная

г. Обнинск 2023 г.

Область применения

Фонд оценочных средств (ФОС) – является обязательным приложением к рабочей программе дисциплины «Машинное обучение и большие данные» и обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущей и промежуточной аттестации по дисциплине.

Цели и задачи фонда оценочных средств

Целью Фонда оценочных средств является установление соответствия уровня подготовки обучающихся требованиям федерального государственного образовательного стандарта.

Для достижения поставленной цели Фондом оценочных средств по дисциплине «Машинное обучение и большие данные» решаются следующие задачи:

- контроль и управление процессом приобретения обучающимися знаний, умений и навыков предусмотренных в рамках данной дисциплины;
- контроль и оценка степени освоения компетенций предусмотренных в рамках данной дисциплины;
- обеспечение соответствия результатов обучения задачам будущей профессиональной деятельности через совершенствование традиционных и внедрение инновационных методов обучения в образовательный процесс в рамках данной дисциплины.

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

1.1. В результате освоения ОП бакалавриата обучающийся должен овладеть следующими результатами обучения по дисциплине:

Код компетенций	Наименование компетенции	Код и наименование индикатора достижения компетенции
УК-3	Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде	3-УК-3 Знать: основные приемы и нормы социального взаимодействия; основные понятия и методы конфликтологии, технологии межличностной и групповой коммуникации в деловом взаимодействии У-УК-3 Уметь: устанавливать и поддерживать контакты, обеспечивающие успешную работу в коллективе; применять основные методы и нормы социального взаимодействия для реализации своей роли и взаимодействия внутри команды В-УК-3 Владеть: простейшими методами и приемами социального взаимодействия и работы в команде
ОПК-4	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	3-ОПК-4 знать технические и программные средства реализации информационных технологий; знать современные программное обеспечение; знать основные методы и средства защиты информации. У-ОПК-4 уметь использовать возможности вычислительной техники, программного обеспечения, средств защиты информации для решения практических задач. В-ОПК-4 владеть навыками использования современных информационных технологий и программного обеспечения для решения задач профессиональной деятельности; владеть навыками соблюдения требований информационной безопасности при использовании современных информационных технологий и программного обеспечения.
ПК-9	Способен внедрять новые методы и средства технического контроля	3-ПК-9 знать справочную документацию по характеристикам используемых материалов, виды возможных дефектов; знать формы и виды документов, используемых при проведении технического контроля. У-ПК-9 уметь планировать потребности в оборудовании, материально технических ресурсах и персонале для реализации технического контроля; уметь разрабатывать все виды операций, входящих в технологический процесс контроля

		<p>параметров и характеристик изделия; уметь составлять схемы контроля параметров и характеристик изделия.</p> <p>В-ПК-9 владеть навыками организации материально технического обеспечения и контроля параметров и характеристик изделия и наладки необходимого контрольно измерительного оборудования.</p>
--	--	---

1.2. Этапы формирования компетенций в процессе освоения ОП бакалавриата

Компоненты компетенций, как правило, формируются при изучении нескольких дисциплин, а также в немалой степени в процессе прохождения практик, НИР и во время самостоятельной работы обучающегося. Выполнение и защита ВКР являются видом учебной деятельности, который завершает процесс формирования компетенций.

Этапы формирования компетенции в процессе освоения дисциплины:

- **начальный** этап – на этом этапе формируются знаниевые и инструментальные основы компетенции, осваиваются основные категории, формируются базовые умения. Студент воспроизводит термины, факты, методы, понятия, принципы и правила; решает учебные задачи по образцу;
- **основной** этап – знания, умения, навыки, обеспечивающие формирование компетенции, значительно возрастают, но еще не достигают итоговых значений. На этом этапе студент осваивает аналитические действия с предметными знаниями по дисциплине, способен самостоятельно решать учебные задачи, внося коррективы в алгоритм действий, осуществляя коррекцию в ходе работы, переносит знания и умения на новые условия;
- **завершающий** этап – на этом этапе студент достигает итоговых показателей по заявленной компетенции, то есть осваивает весь необходимый объем знаний, овладевает всеми умениями и навыками в сфере заявленной компетенции. Он способен использовать эти знания, умения, навыки при решении задач повышенной сложности и в нестандартных условиях.

Этапы формирования компетенций в ходе освоения дисциплины отражаются в тематическом плане (см. РПД).

1.3. Связь между формируемыми компетенциями и формами контроля их освоения

№ п/п	Контролируемые разделы (темы) дисциплины	Индикатор достижения компетенции	Наименование оценочного средства текущей и промежуточной аттестации
Текущая аттестация, 3 курс			
1.	Раздел 1. Стандарт Ворлдскиллс и спецификация стандартов Ворлдскиллс по компетенции «Машинное обучение и большие данные»	УК-3; ОПК-4; ПК-9; В26	Практические задания. Кейс 1.
2.	Раздел 2. Загрузка и подготовка данных для дальнейшей обработки	УК-3; ОПК-4; ПК-9; В26	Практические задания. Кейс 2.
3	Раздел 3. Разработка математического аппарата	УК-3; ОПК-4; ПК-9; В26	Практические задания. Контрольная.
Промежуточная аттестация, 3 курс			

	зачет с оценкой		Демонстрационный экзамен. Тестирование.
--	-----------------	--	--

2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Конечными результатами освоения программы дисциплины являются сформированные когнитивные дескрипторы «знать», «уметь», «владеть», расписанные по отдельным компетенциям, которые приведены в п.1.1. Формирование этих дескрипторов происходит в процессе изучения дисциплины по этапам в рамках различного вида учебных занятий и самостоятельной работы.

Выделяются три уровня сформированности компетенций на каждом этапе: пороговый, продвинутый и высокий.

Уровни	Содержательное описание уровня	Основные признаки выделения уровня	БРС, % освоения	ECTS/Пятибалльная шкала для оценки экзамена/зачета
Высокий <i>Все виды компетенций сформированы на высоком уровне в соответствии с целями и задачами дисциплины</i>	Творческая деятельность	<i>Включает нижестоящий уровень.</i> Студент демонстрирует свободное обладание компетенциями, способен применить их в нестандартных ситуациях: показывает умение самостоятельно принимать решение, решать проблему/задачу теоретического или прикладного характера на основе изученных методов, приемов, технологий	90-100	A/ Отлично/ Зачтено
Продвинутый <i>Все виды компетенций сформированы на продвинутом уровне в соответствии с целями и задачами дисциплины</i>	Применение знаний и умений в более широких контекстах учебной и профессиональной деятельности, нежели по образцу, большей долей самостоятельности и инициативы	<i>Включает нижестоящий уровень.</i> Студент может доказать владение компетенциями: демонстрирует способность собирать, систематизировать, анализировать и грамотно использовать информацию из самостоятельно найденных теоретических источников и иллюстрировать ими теоретические положения или обосновывать практику применения.	85-89	B/ Очень хорошо/ Зачтено
			75-84	C/ Хорошо/ Зачтено
Пороговый <i>Все виды компетенций сформированы на пороговом уровне</i>	Репродуктивная деятельность	Студент демонстрирует владение компетенциями в стандартных ситуациях: излагает в пределах задач курса теоретически и практически контролируемый материал.	65-74	D/Удовлетворительно/ Зачтено
			60-64	E/Посредственно /Зачтено
Ниже порогового	Отсутствие признаков порогового уровня: компетенции не сформированы. Студент не в состоянии продемонстрировать обладание компетенциями в стандартных ситуациях.		0-59	Неудовлетворительно/ Зачтено

Оценивание результатов обучения студентов по дисциплине осуществляется по регламенту текущего контроля и промежуточной аттестации.

Критерии оценивания компетенций на каждом этапе изучения дисциплины для каждого вида оценочного средства и приводятся в п. 4 ФОС. Итоговый уровень сформированности компетенции при изучении дисциплины определяется по таблице. При этом следует понимать, что граница между уровнями для конкретных результатов освоения образовательной программы может смещаться.

Уровень сформированности компетенции	Текущий контроль	Промежуточная аттестация
высокий	высокий	высокий
	<i>продвинутый</i>	<i>высокий</i>
	<i>высокий</i>	<i>продвинутый</i>
продвинутый	<i>пороговый</i>	<i>высокий</i>
	<i>высокий</i>	<i>пороговый</i>
	продвинутый	продвинутый
	<i>продвинутый</i>	<i>пороговый</i>
	<i>пороговый</i>	<i>продвинутый</i>
пороговый	пороговый	пороговый
ниже порогового	пороговый	ниже порогового
	ниже порогового	-

3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

- Итоговая аттестация по дисциплине является интегральным показателем качества теоретических и практических знаний и навыков обучающихся по дисциплине и складывается из оценок, полученных в ходе текущей и промежуточной аттестации.
- Текущая аттестация в семестре проводится с целью обеспечения своевременной обратной связи, для коррекции обучения, активизации самостоятельной работы обучающихся.
- Промежуточная аттестация предназначена для объективного подтверждения и оценивания достигнутых результатов обучения после завершения изучения дисциплины.
- Текущая аттестация осуществляется два раза в семестр:
 - o контрольная точка № 1 (КТ № 1) – выставляется в электронную ведомость не позднее 8 недели учебного семестра. Включает в себя оценку мероприятий текущего контроля аудиторной и самостоятельной работы обучающегося по разделам/темам учебной дисциплины с 1 по 8 неделю учебного семестра.
 - o контрольная точка № 2 (КТ № 2) – выставляется в электронную ведомость не позднее 16 недели учебного семестра. Включает в себя оценку мероприятий текущего контроля аудиторной и самостоятельной работы обучающегося по разделам/темам учебной дисциплины с 9 по 16 неделю учебного семестра.
- Результаты текущей и промежуточной аттестации подводятся по шкале балльно-рейтинговой системы.

Этап рейтинговой системы / Оценочное средство	Неделя	Балл
--	---------------	-------------

		Минимум*	Максимум**
Текущая аттестация	1-16	36 - 60% от максимума	60
Контрольная точка № 1	7-8	18 (60% от 30)	30
<i>Практические задания</i>	1-6	60% от М1	М1
<i>Кейс 1</i>	7-8	60% от М2	М2
Контрольная точка № 2	15-16	18 (60% от 30)	30
<i>Практические задания</i>	9-14	60% от Т1	Т1
<i>Кейс 2</i>	15-16	60% от Т2	Т2
Промежуточная аттестация	-	24 – (60% 40)	40
Зачет с оценкой	-	60% от КР	КР
ИТОГО по дисциплине		60	100

4. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков

ОТДЕЛЕНИЕ ядерной физики и технологий

Направление	<u>«Приборостроение»</u>
Образовательная программа	<u>«Приборы и методы контроля качества и диагностики»</u>
Дисциплина	<u>«Машинное обучение и большие данные»</u>

Кейс-задача 1

по дисциплине «Машинное обучение и большие данные»

Модуль 1.

1.1 ВЫДЕЛЕНИЕ ИТОГОВОГО НАБОРА ПОЛЕЙ ДЛЯ ОДНОЙ ЗАПИСИ Архив obv.xml_205.gz содержит в себе xml файл с исходным набором записей. Каждая запись представлена набором полей. Так же имеется xsd файл, описывающий предоставленный xml файл. Необходимо проанализировать перечисленные поля, их тип, содержание и определить итоговый набор полей для реляционной или нереляционной структуры данных, в которую преобразованные записи будут импортированы

1.2 РАЗБИЕНИЕ ПОЛЯ С МНОЖЕСТВОМ ЗНАЧЕНИЙ НА НЕСКОЛЬКО ПОЛЕЙ

В исходных записях имеются поля, конкатенирующие в себе несколько значений, либо указывающие диапазон. Необходимо выделить такие поля и разбить на несколько других.

1.3 ОПИСАНИЕ СХЕМЫ ДАННЫХ ДЛЯ ИТОГОВОЙ МОДЕЛИ ДАННЫХ Для итогового набора полей необходимо описать xsd файл, сделать это можно по примеру исходного xsd файла.

1.4 СОЗДАНИЕ БАЗЫ ДАННЫХ ДЛЯ ИТОГОВОГО НАБОРА ПОЛЕЙ Необходимо создать реляционную базу данных или любую другую структуру данных, которая позволит сохранить записи, состоящие из итогового набора полей. Так же необходимо предусмотреть разделение на тестовую и обучающую выборки.

1.5 УСТРАНЕНИЕ ДУБЛИРУЮЩИХСЯ И ПУСТЫХ ЗАПИСЕЙ

Из исходных данных необходимо убрать пустые и дублирующие записи.

1.6 ОЧИСТКА ЗНАЧЕНИЙ ОТ ТЭГОВ, СПЕЦ. СИМВОЛОВ

Из значений в исходных полях необходимо убрать все html, xml и любые другие теги. Также значения должны быть очищены от спец. символов.

1.7 ПРЕОБРАЗОВАНИЕ ИЗВЛЕЧЁННЫХ ЗНАЧЕНИЙ К ПОСТРОЕННОЙ МОДЕЛИ (ПАРСИНГ)

Исходные данные необходимо преобразовать к схеме данных для итогового набора полей, разработанной Вами.

1.8 РАЗБИЕНИЕ ПРЕОБРАЗОВАННОЙ ВЫБОРКИ НА ОБУЧАЮЩУЮ И ПРОВЕРОЧНУЮ ВЫБОРКИ

Выборка с преобразованными полями должна быть разбита на 2 части. 1 часть – тестовая выборка, 2 часть – обучающая выборка.

1.9 ВЫГРУЗКА КАЖДОЙ ВЫБОРКИ В СОЗДАННУЮ БАЗУ ДАННЫХ Тестовая и обучающая выборки должны быть импортированы в базу данных или любую другую структуру данных, созданную Вами ранее.

Модуль 2.

2.1 ПОДБОР АЛГОРИТМОВ ВЫЯВЛЕНИЯ ЗАВИСИМОСТЕЙ ПОЛЕЙ ДАННЫХ

Используя программные средства, визуализируйте зависимости полей в данных, загруженных в базу данных ранее. Необходимо произвести выборку данных из базы данных, произвести расчеты зависимостей полей по выбранному алгоритму, визуализировать зависимости выбранным способом.

2.2 ПОДГОТОВКА ОТЧЕТА О ЗАВИСИМОСТИ ДАННЫХ

Подготовьте отчет, содержащий следующие пункты:

- Обоснование выбора алгоритмов выявления зависимостей полей данных
- Выбор способа визуализации зависимостей с обоснованием выбора
- Результаты визуализации

Сохраните отчет в папку с исходными кодами разработанного приложения

Модуль 3.

3.1 РЕАЛИЗАЦИЯ АЛГОРИТМА КЛАССИФИКАЦИИ

Выберите и реализуйте алгоритм для проведения классификации объявлений. Входными параметрами являются все данные, хранящиеся в базе данных, за исключением поля jobname. Данное поле является идентификатором класса. Реализуйте данный алгоритм для классификации новых вакансий.

3.2 РЕАЛИЗАЦИЯ АЛГОРИТМА ОБУЧЕНИЯ

Определите обучающий алгоритм, позволяющий на основе обучающей выборки определять значения поля jobname. Реализуйте выбранный алгоритм.

3.3 ТЕСТИРОВАНИЕ АЛГОРИТМА ОБУЧЕНИЯ

Протестируйте работу обучающего алгоритма. Вычислите процент верного распознавания = (Правильно распознанные вакансии)/(общее количество вакансий)

3.4 ПОДГОТОВКА ОТЧЕТА О РЕЗУЛЬТАТАХ ОБУЧЕНИЯ

Подготовьте отчет, содержащий следующие пункты:

- Обоснование и результаты выбора алгоритма классификации
- Обоснование и результаты выбора алгоритма обучения
- Результаты обучения и тестирования

Сохраните отчет в папку с исходными кодами разработанного приложения

Модуль 4.

4.1 РАЗРАБОТКА ДОКУМЕНТАЦИИ СИСТЕМЫ

В ходе выполнения конкурсного задания вы создавали отчёты по результатам каждой из сессий. Для выполнения данного задания необходимо создать документацию, охватывающую всю вашу выполненную работу. Обоснование выбранных алгоритмов и способов реализации. Скриншоты готового решения и то, что вы посчитаете правильным указать для ознакомления с Вашей работой стороннего наблюдающего.

4.2 ПРЕЗЕНТАЦИЯ РЕЗУЛЬТАТОВ РАБОТЫ

Необходимо создать презентацию, охватывающую все Ваши результаты выполнения тестового задания. В неё должно быть указано ёмкое описание результатов работы с обоснованием выбора того или иного решения. Так же в презентации необходимо предоставить скриншоты результатов своей работы.

Кейс-задача 2

по дисциплине «Машинное обучение и большие данные»

Модуль 1: Предобработка данных и визуализация зависимостей

Исходные файлы:

- | | |
|------------------|--------------------------|
| 1) test_ses.csv | (Тестовый набор данных) |
| 2) train_ses.csv | (Обучающий набор данных) |
| 3) site.pkl | (Словарь сайтов) |
| 4) Модуль 1.pdf | (Инструкция к 1 модулю) |

Результаты работы:

- | | |
|------------------------------------|------------------------------|
| 1) Data.zip | (Предобработанные данные) |
| 2) Report_M1.html, Report_M1.ipynb | (Отчет о проделанной работе) |
| 3) Readme.txt | (Дополнительные комментарии) |

ВВЕДЕНИЕ

На этом экзамене вам предстоит решить задачу определения злоумышленника по его поведению в сети Интернет. По последовательности из десяти веб-сайтов, посещенных подряд одним и тем же человеком, мы будем идентифицировать этого человека. Идея такая: пользователи Интернета по-разному переходят по ссылкам, и это может помочь их идентифицировать (кто-то сначала в почту, потом про футбол почитать, затем новости, социальная сеть, потом, наконец, – работать, кто-то – сразу работать, если это возможно).

Набор данных содержит информацию о сеансах просмотра пользователями, в которых:

- site_i - это идентификаторы сайтов в этом сеансе (в соответствии с словарем site.pkl);
- time_j - это отметки времени посещения соответствующего сайта;
- target - принадлежит ли эта сессия злоумышленнику;

В этом модуле необходимо выполнить подготовку данных к анализу и построению моделей.

1.1 Подготовка обучающей и тестовой выборок

Необходимо выполнить подготовку данных для дальнейшего описательного анализа и построения прогнозных моделей. Следует выполнить загрузку и преобразование всех необходимых данных. Данные необходимо очистить и привести к приемлемому формату

1.2 Работа с разреженным форматом данных

Сформировать мешок сайтов. То есть необходимо создать новые матрицы, в которых строкам будут соответствовать сессии из 10 сайтов, а столбцам – индексы сайтов. На пересечении строки и столбца будет стоять число – количество раз, которое встретился сайт в сессии номер N.

Сериализуйте полученные матрицы для дальнейшего применения и возможного улучшения модели.

1.3 Визуальный анализ данных

Используя программные средства, визуализируйте зависимости атрибутов в наборе данных. Визуализация должна отражать влияние атрибутов на целевую переменную. Приведите интерпретацию полученным результатам.

1.4 Конструирование признаков (Feature Engineering)

Создайте такой признак, который будет представлять собой число вида ГГГГММ от той даты, когда проходила сессия. Например, 201407 - 2014 год и 7 месяц. Таким образом, мы будем учитывать помесечный линейный тренд за весь период предоставленных данных. Добавьте новые признаки, которые на ваш взгляд позволят улучшить качество выбранной модели. Напишите функцию для создания новых признаков и примените ее к исходным данным. Опишите приемы генерации новых данных и результаты. Проведите выбор признаков, т.е. удалите часть признаков, чтобы помочь модели лучше обобщать новые данные ради повышения её точности.

Проведите визуальное исследование полученных признаков и сделайте вывод об их значимости.

1.5 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip

Описание модуля 2:

Исходные файлы:

- | | |
|-----------------|---------------------------------|
| 1) sample.csv | (Пример файла прогнозов) |
| 2) Data.zip | (Результаты предыдущего модуля) |
| 3) Модуль 2.pdf | (Инструкция ко 2 модулю) |

Результаты работы:

- | | |
|------------------------------------|------------------------------|
| 1) Data.zip | (Предобработанные данные) |
| 2) Report_C2.html, Report_C2.ipynb | (Отчет о проделанной работе) |

Readme.txt

ВВЕДЕНИЕ

В этом модуле вы продолжаете работать с данными, подготовленными в предыдущем модуле. На основе предобработанных данных необходимо построить и обучить модель классификации, а также добиться наиболее точной работы построенной модели. Какая-либо работа, обусловленная задачами предыдущего модуля, выполненная в ходе текущего, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

ЗАДАНИЕ

2.1 Сравнение нескольких алгоритмов на сессиях

Выполните разбиение выборки на обучающую и отложенную. Проведите обучение моделей, основанных на различных алгоритмах. Сделайте прогноз в виде предсказанных вероятностей того, что сессия является сессией злоумышленника на отложенной выборке. Выполните оценку моделей разной степени сложности в соответствии с выбранной метрикой, предусмотрите распараллеливание процессов. Выбрать модель, показавшую лучшее качество. Дополните набор данных признаками, полученными в предыдущем модуле, чтобы добиться более точной работы выбранной модели.

2.2 Гиперпараметрическая оптимизация модели

После выбора модели необходимо оптимизировать её под решаемую задачу, настраивая гиперпараметры. Построить для данной модели кривую валидации (зависимость качества классификации от одного из гиперпараметров алгоритма) и кривую обучения (зависимость качества классификации от объема выборки). Выполните оценку данной модели после настройки гиперпараметров.

2.3 Прогноз классификации для тестовой выборки

Сделайте прогноз в виде предсказанных вероятностей отнесения к классу 1 для тестовой выборки с помощью полученной модели, обученной уже на всей обучающей выборке. Запишите ответы в файл в соответствии с шаблоном `sample_sub.csv`

2.4 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам модуля, в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива `Data.zip`

Описание модуля 3:

Исходные файлы:

- 1) Data.zip (Результаты предыдущего модуля)

Результаты работы:

- 1) Data.zip (Результаты модуля)
2) Report.html + Report.ipynb (Отчет о проделанной работе)

ВВЕДЕНИЕ

Вы продолжаете работать с данными, подготовленными в предыдущем модуле. Вам предстоит выполнить регрессионный анализ данных. Модель должна строить предсказания на среднесрочный период. Какая-либо работа, обусловленная задачами предыдущего модуля, выполненная в ходе текущего, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

ЗАДАНИЕ

3.1 Построение регрессионной модели

Для предсказания вероятностей того, что сессия является сессией злоумышленника построить регрессионную модель. Итоговая модель должна быть выбрана из нескольких построенных моделей на алгоритмах различной степени сложности. Для получения оптимальной модели предсказания необходимо выбрать метрику оценки качества работы моделей.

После выбора модели необходимо оптимизировать её под решаемую задачу, настраивая гиперпараметры. Построить для данной модели кривую валидации (зависимость качества от одного из гиперпараметров алгоритма) и кривую обучения (зависимость качества от объема выборки). Выполните оценку данной модели после настройки гиперпараметров

3.2 Визуализация работы регрессионной модели

Визуализируйте предсказания, полученные в результате работы построенной регрессионной модели для нескольких последующих месяцев, начиная с текущего. Визуализация должна наглядно показывать прогнозируемое количество новых случаев заражения для различных стран

3.3 Подготовка отчета

Подготовьте отчет о проделанной работе по итогам Модуля 3 в котором будут представлены результаты, выводы и обоснования выбора по каждому разделу задания. В отчете также опишите содержимое результирующих файлов архива Data.zip

Описание модуля 4:

Исходные файлы:

- | | |
|-------------|---------------------------------|
| 1) Data.zip | (Результаты предыдущего модуля) |
|-------------|---------------------------------|

Результаты работы:

- | | |
|-------------------------------|------------------------------|
| 1) Data.zip | (Результаты модуля) |
| 2) Report.html + Report.ipynb | (Отчет о проделанной работе) |
| 3) Readme.pdf | (Руководство пользователя) |

ВВЕДЕНИЕ

Вы продолжаете работать с данными, подготовленными в предыдущем модуле. Вам предстоит разработать программный продукт – чат-бота. Какая-либо работа, обусловленная задачами предыдущего модуля, выполненная в ходе текущего, оцениваться не будет, поэтому проделывайте её только в случае необходимости.

ЗАДАНИЕ

4.1 Разработка виджета

Вам необходимо разработать программный продукт, который будет выполнять следующие действия:

- позволит пользователю вводить данные в удобном виде;
- выдавать результат предсказания целевой переменной;

осуществлять вывод информации по обработке и анализу данных в отдельный файл.

4.2 Интеграция справки в приложение

Дополнить разработанный программный продукт модулем справочной информацией (help).

4.3 Подготовка руководства пользователя

Подготовьте отчет о проделанной работе по итогам Модуля 4, в котором будут представлены результаты, комментарии к коду, по каждому разделу задания.

Разработайте руководство пользователя для виджета, в котором описать интерфейс и возможности для потенциального пользователя, оформите его в виде файла Readme.pdf.

Кейс-задача 3

по дисциплине «Машинное обучение и большие данные»

Модуль 1 данного Конкурсного задания состоит из следующей документации / файлов:

1. C1_1 DS2018_BDML.pdf (Инструкция к модулю 1)
2. Ск-2бп.wmv (Знакомство с системой УЗК)
3. SOP1.txt, SOP2.txt, SOP3.txt (Исходные данные)
4. Описание ПУЗК_C1_1.doc (Описание технологии автоматизированного УЗК)

ВВЕДЕНИЕ

В этом модуле, Вам предстоит изучить технологию автоматизированного ультразвукового контроля, загрузить данные контроля, предварительно их изучить, очистить от ненужных данных для последующего анализа с учетом длины сварного соединения.

ИНСТРУКЦИЯ УЧАСТНИКУ

К концу этой сессии, у вас должны быть достигнуты следующие результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

1.1 ЗАГРУЗКА ДАННЫХ

Данные загружены. Представлен код по загрузке, отражено в отчете. Результаты контроля очищены от технической информации, отражено в отчете.

1.2 ПРЕДВАРИТЕЛЬНОЕ ИЗУЧЕНИЕ

Представлены распределения переменных (таблица (максимальное, минимальное значение, среднеквадратичное отклонение, медиана, квартили) /графики (значения в координатах и распределений, “ящики с усами”)), отражено в отчете.

1.3 ОПРЕДЕЛЕНИЕ МЕТОДОВ КОНТРОЛЯ

Обоснованно определены методы УЗ контроля (эхо-метод / теневой метод / эхо-контактный) для каждого столбца исходных данных, отражено в отчете.

1.4 ОТЧЕТ

Предоставлен отчет о проделанной работе. Внимание: оценка модуля будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода и результат работы программы. Отчет должен быть предоставлен в папке C[X]_M1, где [X] – номер участника, который должен быть размещен на рабочем столе ноутбука участника. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M1.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M1.HTML(или PDF), где [X] – номер участника, экспортированная тетрадь (или аналог) в статический формат.

Модуль 2.

Модуль 2 данного Конкурсного задания состоит из следующей документации / файлов:

1. C1_2 DS2018_BDML.pdf (Инструкция к модулю 2)
2. SOP1.txt, SOP2.txt, SOP3.txt (Исходные данные)
3. Видеоролик (Знакомство с системой УЗК)
4. Описание ПУЗК_C1_2.doc (Описание технологии автоматизированного УЗК)

ВВЕДЕНИЕ

В этом модуле, Вам предстоит изучить технологию автоматизированного ультразвукового контроля, загрузить данные контроля, предварительно их изучить, очистить от ненужных данных для последующего анализа с учетом длины сварного соединения.

ИНСТРУКЦИЯ УЧАСТНИКУ

К концу этой сессии, у вас должны быть достигнуты следующие результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

2.1 ОПРЕДЕЛЕНИЕ РАССТОЯНИЯ МЕЖДУ СХЕМАМИ ПРОЗВУЧИВАНИЯ

1. Взаимно-корреляционные функции представлены примерами (текст, графики), отражено в отчете. 2. В отчете сделан и описан обоснованный вывод о высоких корреляциях переменных. 3. Определено расстояние между схемами прозвучивания.

2.2 ОПРЕДЕЛЕНИЕ СХЕМ ПРОЗВУЧИВАНИЯ

На основании предыдущего модуля и проведенного корреляционного анализа: 1. Обоснованно определены схемы прозвучивания (хордовая\раздельно-совмещенная) для каждого столбца исходных данных, отражено в отчете. 2. Обоснованно определены схемы прозвучивания (эхо-продольные\эхо-продольные-контактные\эхопоперечные\эхо-поперечные-контактные\теневые) для каждого столбца исходных данных, отражено в отчете. 3. Обоснованно определены схемы прозвучивания (Сторона А\Сторона В) для каждого столбца исходных данных, отражено в отчете. 4. Пересобрана таблица 1 (Описание ПУЗК_C1_M2.doc) в соответствие с номером столбца (такта) в файлах SOP1, SOP2, SOP3

2.3 ПРИВЕДЕНИЕ К ОБЩЕЙ КООРДИНАТЕ

1. Выявлены и обработаны незначащие данные, учтены, отражено в отчете. 2. Выявлены, обработаны и обоснованы смещения в данных (данные приведены к общей координате), отражено в отчете.

2.4 ОТЧЕТ

Предоставлен отчет о проделанной работе. Внимание: оценка сессии будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода, и результат работы программы. Отчет должен быть предоставлен в папке C[X]_M2, где [X] – номер участника, который должен быть размещен на рабочем столе ноутбука

участника. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M2.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M1.HTML(или PDF), где [X] – номер участника, экспортированная тетрадь (или аналог) в статический формат

Модуль 3.

Модуль 3 данного Конкурсного задания состоит из следующей документации / файлов:

1. C2_3_DS2018_BDML.pdf (Инструкция к модулю 3)
2. Видеоролик (Знакомство с системой ПУЗК)
3. SOP1.txt, SOP2.txt, SOP3.txt (Исходные данные)
4. target1.csv, target2.csv, target3.csv (Целевая переменная)
5. Описание ПУЗК_C2_M3.doc (Описание технологии полуавтоматизированного УЗК)

ВВЕДЕНИЕ

В этом модуле Вам предстоит выбрать переменные из исходного набора данных, при необходимости создать новые переменные и приступить к подбору алгоритма для определения целевой переменной.

ИНСТРУКЦИЯ УЧАСТНИКУ

К концу этой сессии, у вас должны быть достигнуты следующие результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

3.1 ВЫБОР ПЕРЕМЕННЫХ ИЗ ИСХОДНОГО НАБОРА ДАННЫХ: ОБОСНОВАНИЕ (ЛОГИКА, АНАЛИЗ ДАННЫХ)

Представлена (в текстовом или графическом виде) логика выбора переменных для построения модели, отражено в отчете. Переменные обоснованы исходя из предметной области, природы данных, отражено в отчете.

3.2 ВЫБОР И ПОСТРОЕНИЕ НОВЫХ ПЕРЕМЕННЫХ; ТЕСТИРОВАНИЕ ЗНАЧИМОСТИ ВЫБРАННЫХ ПЕРЕМЕННЫХ

Описаны необходимые шаги по обработке данных для получения переменных, отражено в отчете. Написана функция для получения переменных, отражено в отчете. Представлена и оценена значимость выбранных переменных, отражено в отчете.

3.3 АРГУМЕНТИРОВАННЫЙ ВЫБОР АЛГОРИТМА ДЛЯ КЛАСТЕРИЗАЦИИ

Алгоритм выбран на основе сравнения эффективности не менее двух вариантов на существующих данных, отражено в отчете. Описана в текстовом виде логика выбора алгоритма, отражено в отчете.

3.4 ОТЧЕТ

Предоставлен отчет о проделанной работе. Внимание: оценка сессии будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода, и результат работы программы. Отчет должен быть предоставлен в папке C[X]_M3, где [X] – номер участника, который должен быть размещен на рабочем столе ноутбука участника. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M3.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M1.HTML(или PDF), где [X] – номер участника, экспортированная тетрадь (или аналог) в статический формат

Модуль 4.

Модуль 4 данного Конкурсного задания состоит из следующей документации / файлов:

1. C2_4_DS2018_BDML (Инструкция к модулю 4)
2. SOP1.csv, SOP2.csv, SOP3.csv, (Исходные данные)
3. SOP1.txt, SOP2.txt, SOP3.txt (Исходные данные)
4. Папка “Results” с результатами контроля (Публичная контрольная выборка)
5. Папка “Targets” с результатами расшифровки (Публичная контрольная выборка)

ВВЕДЕНИЕ

В этом модуле Вам необходимо разработать модель, произвести разбиение данных на обучающую и проверочную выборку, а также, сделать предсказание целевой переменной на основе предоставленной публичной контрольной выборки.

ИНСТРУКЦИЯ УЧАСТНИКУ

К концу этой сессии, у вас должны быть достигнуты следующие результаты.

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

4.1 РАЗБИЕНИЕ ДАННЫХ НА ОБУЧАЮЩУЮ И ПРОВЕРОЧНУЮ ВЫБОРКУ

Подготовлены файлы обучающей и проверочной выборки. Либо представлен алгоритм по формированию этих выборок, отражено в отчете

4.2 ВИЗУАЛЬНОЕ ОБОСНОВАНИЕ ВЫБОРА ПЕРЕМЕННЫХ/ РАБОТОСПОСОБНОСТИ МОДЕЛИ

Представлены графики влияния переменных на целевую переменную. Представлен график матрицы ошибок, отражено в отчете

4.3 ПРОВЕРКА ТОЧНОСТИ НА ПУБЛИЧНОЙ КОНТРОЛЬНОЙ ВЫБОРКЕ

Предоставлен набор данных в установленном формате: идентификатор строки, предсказание целевой переменной. Предоставлен и документирован код создающий требуемый набор данных. В отчете должна быть представлена оценка точности модели на обучающей выборке

4.4 ОТЧЕТ

Предоставлен отчет о проделанной работе. Внимание: оценка модуля будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода и результат работы программы. Отчёт должен быть предоставлен в папке C[X]_M4, где [X] – номер команды, которая должна быть размещена на рабочем столе Windows ноутбука одного из участников. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M4.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M4.HTML(или PDF), где [X] – номер команды, экспортированная тетрадь (или аналог) в статический формат.

Модуль 5.

Модуль 5 данного Конкурсного задания состоит из следующей документации / файлов:

1. C3_5_DS2018_BDML (Инструкция к четвертому модулю)
2. SOP1.csv, SOP2.csv, SOP3.csv, (Исходные данные)
3. sop1_target.csv, sop2_target.csv, sop3_target.csv (Исходные данные целевой переменной)
4. Папка "Results" с результатами контроля (Публичная контрольная выборка)
5. Папка "Targets" с результатами расшифровки (Публичная контрольная выборка)

ВВЕДЕНИЕ

В этом модуле Вам предстоит разработать удобный графический интерфейс с возможностью ввести результаты контроля и сделать предсказание целевой переменной.

ИНСТРУКЦИЯ УЧАСТНИКУ

К концу этой сессии, у вас должны быть достигнуты следующие результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

5.1 РАЗРАБОТКА ИНТЕРФЕЙСА

Разработан интерфейс с возможностью запуска. Интерфейс предоставляет пользователю возможность обмена информацией и получения результата. Разработана короткая инструкция по запуску и эксплуатации интерфейса (файл с названием Инструкция в формате docx или txt).

5.2 УДОБСТВО РАБОТЫ И ДОКУМЕНТИРОВАНИЕ

Интерфейс позволяет пользователю вводить данные в удобном виде. Результатом работы интерфейса являются корректные предсказания целевых переменных. Исходный код прокомментирован и внесен в отчет.

5.3 ОТЧЕТ

Предоставлен отчет о проделанной работе. Внимание: оценка сессии будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода, и результат работы программы. Отчет должен быть предоставлен в папке C[X]_M5, где [X] – номер участника, который должен быть размещен на рабочем столе ноутбука участника. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M5.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M5.HTML(или PDF), где [X] – номер участника, экспортированная тетрадь (или аналог) в статический формат

Модуль 6.

Модуль 6 данного Конкурсного задания состоит из следующей документации / файлов:

1. C3_6_DS2018_BDML (Инструкция к шестому модулю)
2. Папка “Results” с результатами контроля (Исходные данные)
3. Папка с результатами расшифровки (Результаты расшифровки)
4. Папка “Validate” (Приватная выборка)

ВВЕДЕНИЕ

В этом модуле Вам предстоит разработать презентацию результатов выполнения конкурсного задания, сделать устный доклад, рассказать о перспективах улучшения модели. Так же, сделать предсказание целевой переменной на основе предоставленной приватной контрольной выборки.

ИНСТРУКЦИЯ УЧАСТНИКУ ВНИМАНИЕ:

В конце модуля будет выдана приватная контрольная выборка. Формат приватной выборки соответствует публичной выборке. Очень важно, чтобы Ваше решение справилось с предсказанием целевой переменной при запуске кода, без ошибок кода. Участникам будет дана возможность удостовериться в завершении предсказания целевой переменной без ошибок кода (на исправление ошибок кода максимум 15 минут, после обнаружения ошибки). К концу этой сессии, у вас должны быть достигнуты следующие результаты:

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

6.1 РАЗРАБОТКА ПРЕЗЕНТАЦИИ РЕЗУЛЬТАТОВ КОНКУРСНОГО ЗАДАНИЯ

Подготовлена презентация (текстовый или графический файл, описывающий результаты работы и процесс достижения результата по всем модулям в целом). Разработана документация.

6.2 ДОКЛАД ПО ПРЕЗЕНТАЦИИ

Представлен устный доклад по презентации (тайминг – 5 минут)

6.3 ПЕРСПЕКТИВЫ УЛУЧШЕНИЯ МОДЕЛИ

Представлены перспективы улучшения модели в рамках доклада, документации и презентации.

6.4 ТОЧНОСТЬ НА ПРИВАТНОЙ КОНТРОЛЬНОЙ ВЫБОРКЕ

Предоставлен набор данных в установленном формате: предсказание целевой переменной. Предоставлен и документирован код создающий требуемый набор данных. 6.5 ОТЧЕТ
Предоставлен отчет о проделанной работе. Внимание: оценка сессии будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода, и результат работы программы. Отчет должен быть предоставлен в папке C[X]_M6, где [X] – номер участника, который должен быть размещен на рабочем столе ноутбука участника. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. В папке 2 файла – Jupyter Notebook C[X]_M6.ipynb (или аналог – с возможностью запустить и исполнить участки кода), C[X]_M6.HTML(или PDF), где [X] – номер участника, экспортированная тетрадь (или аналог) в статический формат

Критерии и шкала оценивания

Оценка	Критерии оценки
Отлично с 25 до 30 баллов	Студент демонстрирует свободное обладание компетенциями, способен применить их в нестандартных ситуациях: показывает умение самостоятельно принимать решение, решать проблему/задачу теоретического или прикладного характера на основе изученных методов, приемов, технологий
Хорошо с 20 до 25 баллов	<i>Включает нижестоящий уровень.</i> Студент может доказать владение компетенциями: демонстрирует способность собирать, систематизировать, анализировать и грамотно использовать информацию из самостоятельно найденных теоретических источников и иллюстрировать ими теоретические положения или обосновывать практику применения.
Удовлетворительно с 15 до 20 баллов	Студент демонстрирует владение компетенциями в стандартных ситуациях: излагает в пределах задач курса теоретически и практически контролируемый материал.
Неудовлетворительно с 0 до 15 баллов	Отсутствие признаков порогового уровня: компетенции не сформированы. Студент не в состоянии продемонстрировать обладание компетенциями в стандартных ситуациях.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
Обнинский институт атомной энергетики –
филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(ИАТЭ НИЯУ МИФИ)

ОТДЕЛЕНИЕ ядерной физики и технологий

Направление	<u>«Приборостроение»</u>
Образовательная программа	<u>«Приборы и методы контроля качества и диагностики»</u>
Дисциплина	<u>«Машинное обучение и большие данные»</u>

Комплект тестовых заданий

Вопрос № 1 Большие данные – это:

- А) Данные объемом более 1Тб
- Б) Данные объемом более 10Тб
- В) Данные объемом более 100Тб
- Г) Нет ограничений на минимальный объем**

Вопрос № 2 Что, из ниже перечисленного, относится к обучающей выборке?

- А) Классификация данных
- Б) Объекты с известными ответами**
- В) Алгоритм, решающий функцию
- Г) Все перечисленное

Вопрос № 3 Что называют данными в машинном обучении?

- А) Матрицы**
- Б) Объекты
- В) Признаки
- Г) Алгоритмы

Вопрос № 4 Признак F1 может принимать значения A, B, C, D, E и F и отображает оценку ученика колледжа. Какое из следующих утверждений верно?

- A) F1 - номинальная переменная
- Б) F1 - ординальная переменная**
- В) F1 не относится ни к одному из этих типов переменных
- Г) F1 можно отнести к обоим типам

Вопрос № 5 Азиатский пользователь жалуется, что модель распознавания лиц вашей компании неправильно идентифицирует выражение его лица. Что вы должны делать?

- A) Включите азиатские лица в свои тестовые данные и переобучите свою модель
- Б) Переобучите свою модель с помощью обновленных значений гиперпараметров
- В) Переобучите свою модель с меньшими размерами выборки
- Г) Включите азиатские лица в свои тренировочные данные и переобучите свою модель**

Вопрос № 6 Укажите показатели вариации

- A) Мода и медиана
- Б) Стандартное отклонение и дисперсия**
- В) Темп роста и прироста
- Г) Все перечисленные

Вопрос № 7 Выборка может быть: а) случайная, б) механическая, в) типическая, г) серийная, д) техническая

- А) а,б,в,г**
- Б) а,б,в,д
- В) б,в,г,д
- Г) а,б,в,г,д

Вопрос № 8 Ряд динамики характеризует:

- A) Структуру совокупности по какому-то признаку

Б) Изменение характеристик совокупности во времени

В) Определенное значение признака в совокупности

Г) Величину показателя на определенную дату или за определенный период

Вопрос № 9 Когда говорят о нейронных сетях и машинном обучении, часто упоминают закон Мура. В чем его суть?

А) Каждое следующее поколение компьютеров работает в 2,5 раза быстрее

Б) Не следует множить сущее без необходимости

В) 20% усилий дают 80% результата, а остальные 80% усилий — лишь 20% результата

Г) Если все слова языка или длинного текста упорядочить по убыванию частоты их использования, то частота n -го слова в таком списке окажется приблизительно обратной пропорциональной его порядковому номеру n

Вопрос № 10 Медиана - ...

А) Среднее значение признака в ряду распределения

Б) Наиболее часто встречающееся значение признака в данном ряду

В) Значение признака, делящее совокупность на две равные части

Г) Наиболее редко встречающееся значение признака в данном ряду

Вопрос № 11 22 – средняя величина признака; – 26 % – коэффициент вариации признака Дисперсия признака (точность до 0,1) = _____?

А) 28

Б) 35,6

В) 27,8

Г) 32,7

Вопрос № 12 Что такое вариация?

А) Изменение массовых явлений во времени

Б) Изменение структуры статистической совокупности в пространстве

В) Изменение значений признака во времени и в пространстве

Г) Изменение состава совокупности

Вопрос № 13 Что из перечисленного включает этап анализа и подготовки данных?

А) Сбор данных

Б) Нормализация данных

В) Выбор алгоритма для моделирования

Г) Конструирование признаков

Вопрос № 14 Общими мерами вариабельности данных являются:

А) Дисперсия

Б) Мода и медиана

В) Стандартное отклонение

Г) Межквартильный размах

Вопрос № 15 В наборе данных, с которым Вам предстоит работать, много пропущенных значений. Что из перечисленного поможет Вам лучше решить проблему?

А) Использовать генератор случайных значений для заполнения пропусков

Б) Заменить отсутствующие значения средним значением по параметру

В) Удалить записи, содержащие пустые значения

Г) Восстановить данные, используя специальные алгоритмы

Вопрос № 16 Диаграмма, построенная из набора числовых данных, в центре которой находится медиана, по сторонам ящика – квартили, далее максимальные и минимальные значения

А) Диаграмма рассеяния

Б) Черный ящик

В) Ящик с усами

Г) Диаграмма Ганта

Вопрос № 17 Некоторые признаки иногда не дают никакого вклада в улучшение качества работы алгоритма, а порой бывают даже вредными. Поэтому используют фильтрацию, отбрасывая ненужные признаки, с учётом некоторого критерия. Такая процедура называется ...

- А) Центрирование
- Б) Масштабирование и нормализация
- В) Отбор признаков**
- Г) One-hot-encoding

Вопрос № 18 Что называют синтезом признаков?

- А) Описание искомого алгоритма как суперпозиции некоторых элементарных функций
- Б) Поиск преобразования исходящего пространства признаков в новое пространство существенно меньшей размерности**
- В) Когда по мере увеличения числа используемых признаков средняя ошибка на обучающей выборке монотонно убывает
- Г) Когда средняя оценка на независимых контрольных данных сначала уменьшается, затем проходит через точку минимума и далее только возрастает

Вопрос № 19 Алгоритмы классификации обычно имеют параметры, которые необходимо установить до начала процесса обучения. Эти параметры известны как _____.

- А) Настраиваемые решетки
- Б) Контролируемые параметры
- В) Гиперпараметры**
- Г) Число степеней свободы модели

Вопрос № 20 Анализ тесноты и направления связей двух признаков осуществляется на основе

- А) Парного коэффициента корреляции**
- Б) Коэффициента детерминации
- В) Множественного коэффициента корреляции

Г) Коэффициента вариации

Зачтено	Количество верных ответов в интервале: 50-100%
Незачтено	Количество верных ответов в интервале: 0-50%

Программу составили:

_____ к.т.н., доцент Белоусов Павел Анатольевич

_____ преподаватель отделения ЯФиТ, Распопов Дмитрий Алексеевич

Рецензент (ы):

_____ д.ф.-м.н., профессор Старков Сергей Олегович

_____ к.т.н., доцент Мирзеабасов Олег Ахметбекович