

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

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

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.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Машинное обучение и большие данные» является ознакомление студентов с основами науки о данных и принципами работы искусственного интеллекта, формирование умения самостоятельно разбираться в имеющихся концепциях, методах, моделях и технологиях машинного обучения в плане реализации эффективных интеллектуальных систем и применять их для решения прикладных задач.

Задачи освоения дисциплины. Учебная дисциплина направлена на совершенствование и получение новой компетенции, необходимой для профессиональной деятельности, и повышение профессионального уровня с учетом спецификации стандарта Ворлдскиллс по компетенции «Машинное обучение и большие данные».

В результате изучения дисциплины происходит подготовка студентов к демонстрационному экзамену по стандартам WorldSkills по компетенции “Машинное обучение и большие данные”.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ (далее – ОП) БАКАЛАВРИАТА

Дисциплина реализуется в рамках факультативной части и относится к профессиональному модулю в рамках апробации обучения студентов вузов по компетенции FutureSkills «Машинное обучение и большие данные».

Для освоения дисциплины необходимы компетенции, сформированные в рамках изучения следующих дисциплин и/или практик (при необходимости): информатика, теория вероятности, математический анализ, линейная алгебра, аналитическая геометрия.

Дисциплины и/или практики, для которых освоение данной дисциплины необходимо как предшествующее: методы и алгоритмы технической диагностики АЭС, компьютерные технологии в технической диагностике.

Дисциплина изучается на 3 курс.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

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	З-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 11 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	З-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	З-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 параметров и характеристик изделия; уметь составлять схемы контроля параметров и характеристик изделия. В-ПК-9 владеть навыками организации материально технического обеспечения и контроля параметров и характеристик изделия и наладки необходимого контрольно измерительного оборудования.

4. ВОСПИТАТЕЛЬНЫЙ ПОТЕНЦИАЛ ДИСЦИПЛИНЫ

Направления/цели воспитания	Задачи воспитания (код)	Воспитательный потенциал дисциплин
Гражданское и патриотическое воспитание	Формирование ответственной позиции по применению ядерных технологий в свете сохранения окружающей среды для будущих поколений (В26)	1. Использование воспитательного потенциала дисциплины «Машинное обучение и большие данные».

5. ОБЪЕМ ДИСЦИПЛИНЫ В ЗАЧЕТНЫХ ЕДИНИЦАХ С УКАЗАНИЕМ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ЧАСОВ, ВЫДЕЛЕННЫХ НА КОНТАКТНУЮ РАБОТУ ОБУЧАЮЩИХСЯ С ПРЕПОДАВАТЕЛЕМ (ПО ВИДАМ ЗАНЯТИЙ) И НА САМОСТОЯТЕЛЬНУЮ РАБОТУ ОБУЧАЮЩИХСЯ

Вид работы	Количество часов на вид работы:
Контактная работа обучающихся с преподавателем	
Аудиторные занятия (всего)	8
В том числе:	
<i>лекции</i>	2
<i>практические занятия</i>	4
<i>лабораторные занятия</i>	2
Промежуточная аттестация	
В том числе:	
<i>зачет</i>	
<i>зачет с оценкой</i>	+
<i>экзамен</i>	
Самостоятельная работа обучающихся	
Самостоятельная работа обучающихся	28
Всего (часы):	36
Всего (зачетные единицы):	1

6. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ, СТРУКТУРИРОВАННОЕ ПО ТЕМАМ (РАЗДЕЛАМ) С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ЧАСОВ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ

6.1. Разделы дисциплины и трудоемкость по видам учебных занятий (в академических часах)

Неделя	Наименование раздела / темы дисциплины	Виды учебной работы				
		Лек	Пр	Лаб	Внеауд	СРО
1	1. Стандарт Ворлдскиллс и спецификация стандартов Ворлдскиллс по компетенции «Машинное обучение и большие данные»	1	2	1		5
1	1.1. Описание компетенции «машинное обучение и большие данные». Спецификация стандартов Worldskills по компетенции «Машинное обучение и большие данные».	1	2	1		9

2-10	2. Загрузка и подготовка данных для дальнейшей обработки	0.5	1	0.5		14
2-8	2.1. Анализ данных	0.25	0.5	0.25		5
9	2.2 Нормализация данных					5
10	2.3 Визуализация зависимостей	0.25	0.5	0.25		4
11-16	3. Разработка математического аппарата	0.5	1	0.5		9
11-12	Предобработка и подготовка данных для обучения модели					3
13-14	Разработка математического аппарата для решения задачи классификации	0.25	0.5	0.25		3
15-16	Разработка математического аппарата для решения задачи регрессии	0.25	0.5	0.25		3
	Итого за 3 курс:	2	4	2		28
	Всего:	2	4	2		28

Прим.: Лек – лекции, Пр – практические занятия / семинары, Лаб – лабораторные занятия, Внеауд – внеаудиторная контактная работа, СРО – самостоятельная работа.

6.2. Содержание дисциплины, структурированное по разделам (темам)

Лекционный курс

Неделя	Наименование раздела / темы дисциплины	Содержание
1	1. Стандарт Ворлдскиллс и спецификация стандартов Ворлдскиллс по компетенции «Машинное обучение и большие данные»	
1	1.1. Описание компетенции «машинное обучение и большие данные». Спецификация стандартов Worldskills по компетенции «Машинное обучение и большие данные».	Представление и разбор технического описания компетенции «Машинное обучение и большие данные». Разбор стандартов Ворлдскиллс.
2-9	2. Загрузка и подготовка данных для дальнейшей обработки	
2-8	2.1. Анализ данных	Анализ полей, их типа, содержания, определение итогового набора полей для базы данных, в которую преобразованные записи будут импортированы для разных по объёму и предметной области наборов входных данных.
11-16	3. Разработка математического аппарата	
13-14	Разработка математического аппарата для решения задачи классификации	Введение в искусственный интеллект. Типы задач. Типы нейронных сетей. Сверточные нейронные сети. Рекуррентные нейронные сети.
15-16	Разработка математического аппарата для решения задачи регрессии	Генетические алгоритмы.

Практические/семинарские занятия

Неделя	Наименование раздела / темы дисциплины	Содержание
2-10	2. Загрузка и подготовка данных для дальнейшей обработки	
2-8	2.1. Анализ данных	Анализ полей, их типа, содержания, определение итогового набора полей для базы данных, в которую преобразованные записи будут импортированы для разных по объёму и предметной области наборов входных данных.
9	2.2 Нормализация данных	Нормализация полей, конкатенирующих в себе несколько значений, либо указывающих диапазон, и, при необходимости, числовых полей для проанализированных и отобранных на предыдущем этапе разных по объёму и предметной области наборов входных данных.
10	2.3 Визуализация зависимостей	Построение диаграмм рассеяния, гистограмм, ящиков с усами, корреляционных матриц и тепловых карт.
11-16	3. Разработка математического аппарата	
11-12	Предобработка и подготовка данных для обучения модели	Подготовка репозитория, парсинг, разбиение выборки на обучающую и тестирующую для проанализированных и нормализованных на предыдущем этапе разных по объёму и предметной области наборов входных данных.
13-14	Разработка математического аппарата для решения задачи классификации	Построение математической модели и применение математических методов в решении задач машинного обучения на проанализированных на предыдущем этапе разных по объёму и предметной области наборов входных данных с использованием современных программных продуктов для построения математических моделей.
15-16	Разработка математического аппарата для решения задачи регрессии	Построение математической модели и применение математических методов в решении задач машинного обучения на проанализированных на предыдущем этапе разных по объёму и предметной области наборов входных данных с использованием современных программных продуктов для построения математических моделей. Создание презентации по результатам работы.

7. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1. П.А. Белоусов, О.В. Марухина, А.О. Скоморохов, С.О. Старков, О.А. Мирзеабасов, Д.А. Распопов, *Машинное обучение и большие данные, Учебное пособие, – СПб.: ГУАП, 2021. – 119 с.*

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

8.1. 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.

8.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущей и промежуточной аттестации по дисциплине.

Оценочные средства приведены в Приложении «Фонд оценочных средств».

8.3. 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:
 - 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.

- 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

8.4. Шкала оценки образовательных достижений

Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет сумму баллов, заработанных обучающимся при выполнении заданий в рамках текущей и промежуточной аттестации

Сумма баллов	Оценка по 4-х балльной шкале	Оценка ECTS	Требования к уровню освоения учебной дисциплины
90-100	5- «отлично»/ «зачтено»	A	Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы
85-89	4 - «хорошо»/ «зачтено»	B	Оценка «хорошо» выставляется обучающемуся, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос
75-84		C	
70--74		D	
65-69	3 - «удовлетворительно»/ «зачтено»	D	Оценка «удовлетворительно» выставляется обучающемуся, если

60-64		E	он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала
0-59	2 - «неудовлетворительно»/ «не зачтено»	F	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится обучающимся, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине

9. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

а) основная учебная литература:

1. Рашка, С. Python и машинное обучение [Электронный ресурс] : руководство / С. Рашка ; пер. с англ. Логунова А.В.. — Электрон. дан. — Москва: ДМК Пресс, 2017. — 418 с. — Режим доступа: <https://e.lanbook.com/book/100905>. — Загл. с экрана.

2. Hastie, T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, - 2009. <https://web.stanford.edu/~hastie/Papers/ESLII.pdf>

б) дополнительная учебная литература:

1. Северенс, Ч. Введение в программирование на Python / Ч. Северенс. - 2-е изд., испр. - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 231 с. : схем., ил. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=429184>

10. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ» (ДАЛЕЕ - СЕТЬ «ИНТЕРНЕТ»), НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1. Официальный сайт Python: <https://www.python.org/>
2. Библиотеки Python: <http://www.numpy.org/>, <http://matplotlib.org/>, <http://scikit-learn.org/stable/>, <http://pandas.pydata.org/>

11. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1. В пятом и шестом семестре студенты познакомились с компьютерными технологиями, языками программирования и основными инструментами для анализа и обработки данных. Студентами получены первичные знания и практические навыки создания математических моделей и алгоритмов для решения задач машинного обучения.

2. Главной целью практических занятий является закрепление полученных в пятом семестре знаний и умений, Это достигается практикой использования современных методов и алгоритмов машинного обучения для решения задач классификации и регрессии на различных наборах исторических данных.

3. В ходе освоения дисциплины под руководством преподавателей студенты выполняют практические задания, изучают новые модели и алгоритмы машинного обучения, готовятся к сдаче демонстрационного экзамена. Для подготовки к демонстрационному экзамену, рекомендуется решить данные кейсы.

1. Разработка системы классификации объявлений.
2. Решение задачи определения злоумышленника по его поведению в сети Интернет.
3. Разработка системы автоматического поиска дефектов в сварных швах трубопроводов промышленных предприятий.

Пример одного из кейсов представлен ниже в рекомендациях. Описание других кейсов находится в приложении ФОС.

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 емкое описание результатов работы с обоснованием выбора того или иного решения. Также в презентации необходимо предоставить скриншоты результатов своей работы.

Проверка, защита работ проектов.

Проверка работ студентов включает в себя следующие этапы.

- а) проверка правильности написания и оформления отчета;
- б) проверка правильности написания кода (наличие и запуск кода);
- в) проверка правильности полученных результатов по работе.

Студент должен выполнить тесты на знание пройденного материала.

Проверяя задания, преподаватель проводит собеседование со студентом по темам. Студенту задаются вопросы, которые помогут выявить степень усвоения учебного материала, закрепить пройденный материал.

Студент обязан:

1. Посещать регулярно практические занятия, выполнять все текущие задания по изучаемой теме.
2. Пройти аттестацию по всем разделам дисциплины.
3. В результате освоения дисциплины выполнить зачетную работу (демонстрационный экзамен).

Для аттестации по разделам и допуску к зачету студенту необходимо получить не менее 60 баллов суммарно по всем разделам. Все практические работы должны быть выполнены студентом и защищены.

Все практические задания должны быть выполнены студентом самостоятельно и сданы преподавателю.

12. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ, ВКЛЮЧАЯ ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ (ПРИ НЕОБХОДИМОСТИ)

Использование информационных технологий при осуществлении образовательного процесса по дисциплине осуществляется в соответствии с утвержденным Положением об Электронной информационно-образовательной среде ИАТЭ НИЯУ МИФИ.

Электронная система управления обучением (LMS) используется для реализации образовательных программ при очном, дистанционном и смешанном режиме обучения. Система реализует следующие основные функции:

- 1) Создание и управление классами,
- 2) Создание курсов,
- 3) Организация записи учащихся на курс,
- 4) Предоставление доступа к учебным материалам для учащихся,
- 5) Публикация заданий для учеников,
- 6) Оценка заданий учащихся, проведение тестов и отслеживание прогресса обучения,
- 7) Организация взаимодействия участников образовательного процесса.

Система интегрируется с дополнительными сервисами, обеспечивающими возможность использования таких функций как рабочий календарь, видео связь, многопользовательское редактирование документов, создание форм опросников, интерактивная доска для рисования. Авторизация пользователей в системе осуществляется посредством корпоративных аккаунтов, привязанных к домену oiate.ru.

12.1. Перечень информационных технологий

При осуществлении образовательного процесса по дисциплине используются следующие информационные технологии:

- проведение лекций и практических занятий с использованием слайд-презентаций;
- использование текстового редактора Microsoft Word;
- использование табличного редактора Microsoft Excel;
- использование текстового редактора NoteBook (Блокнот);
- использование компьютерного тестирования;
- организация взаимодействия с обучающимися посредством электронной почты и ЭИОС.

12.2. Перечень программного обеспечения

1. Текстовый редактор Microsoft Word;
2. Табличный редактор Microsoft Excel;
3. Редактор презентаций Microsoft PowerPoint;
4. Текстовый редактор NoteBook (Блокнот);
5. Браузеры: Google Chrome, Internet Explorer, Yandex, Mozilla Firefox, Opera.
6. Локальная компьютерная сеть и глобальная сеть Интернет.
7. Дистрибутив языков программирования Python и R – Anaconda

8. Инструмент Python–разработчика JetBrains PyCharm
9. Инструмент для анализа данных Jupyter Notebook
10. Python IDE – интерпретатор Python.

12.3. Перечень информационных справочных систем

Доступ к электронным библиотечным ресурсам и электронной библиотечной системе (ЭБС) осуществляется посредством специальных разделов на официальном сайте ИАТЭ НИЯУ МИФИ. Обеспечен доступ к электронным каталогам библиотеки ИАТЭ НИЯУ МИФИ, а также электронным образовательным ресурсам (ЭИОС), сформированным на основании прямых договоров с правообладателями учебной и учебно-методической литературы, методических пособий:

- 1) Информационные ресурсы Сети Консультант Плюс, www.consultant.ru (информация нормативно-правового характера на основе современных компьютерных и телекоммуникационных технологий);
- 2) Электронно-библиотечная система НИЯУ МИФИ, http://libcatalog.mephi.ru/cgi/irbis64r/cgiirbis_64.exe?C21COM=F&I21DBN=BOOK&Z21ID=&P21DBN=BOOK;
- 3) ЭБС «Издательства Лань», <https://e.lanbook.com/>;
- 4) Электронно-библиотечная система BOOK.ru, www.book.ru;
- 5) Базы данных «Электронно-библиотечная система elibrary» (ЭБС elibrary);
- 6) Базовая версия ЭБС IPRbooks, www.iprbooks.ru;
- 7) Базы данных «Электронная библиотека технического ВУЗа» www.studentlibrary.ru;
- 8) Электронно-библиотечная система «Айбукс.py/ibooks.ru»,
- 9) <http://ibooks.ru/home.php?routine=bookshelf>
- 10) Электронно-библиотечная система «ЭБС ЮРАЙТ», <http://urait.ru/>.

13. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

14. ИНЫЕ СВЕДЕНИЯ И (ИЛИ) МАТЕРИАЛЫ

14.1. Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине

№ пп	Наименование темы дисциплины	Вид занятий (лекция, семинары, практические занятия)	Количество ак. ч.	Наименование активных и интерактивных форм проведения занятий

1	Стандарт Ворлдскиллс и спецификация стандартов Ворлдскиллс по компетенции «Машинное обучение и большие данные»	Лекция	1	Круглый стол, дискуссия, дебаты как форма консультирования студентов.
2	Загрузка и подготовка данных для дальнейшей обработки	практические занятия	2	Мозговой штурм, case-study (коллективный анализ конкретных ситуаций, ситуационный анализ) при поиске вариантов решения задач, сформулированных в проектных заданиях.
3	Разработка математического аппарата	практические занятия	2	Мастер классы, тренинги и симуляции, которые организуют студенты.

14.2. Формы организации самостоятельной работы обучающихся (темы, выносимые для самостоятельного изучения; вопросы для самоконтроля; типовые задания для самопроверки)

Список вопросов для самостоятельной работы

1. Постановка задач обучения по прецедентам.
2. Типы шкал: бинарные, номинальные, порядковые, количественные.
3. Типы задач: классификация, регрессия, прогнозирование, кластеризация. Примеры прикладных задач.
4. Постановка задачи кластеризации. Примеры прикладных задач.
5. Типы кластерных структур. Графовые алгоритмы кластеризации. Выделение связанных компонент.
6. Кратчайший незамкнутый путь. Алгоритм ФОРЭЛ.
7. Функционалы качества кластеризации
8. Статистические алгоритмы: EM-алгоритм и Алгоритм k средних (k-means).
9. Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM
10. Самоорганизующаяся карта Кохонена. Применение для визуального анализа данных.
11. Искусство интерпретации карт Кохонена. Сети встречного распространения, их применение для кусочнопостоянной и гладкой аппроксимации функций
12. Агломеративная кластеризация, Алгоритм Ланса-Вильямса и его частные случаи.

13. Алгоритм построения дендрограммы. Определение числа кластеров. Свойства сжатия/растяжения, монотонности и редуцируемости
14. Метод ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля.
15. Метод окна Парзена.
16. Метрические методы классификации в задаче восстановления регрессии. Обнаружение выбросов.
17. Понятия закономерности и информативности. Понятие логической закономерности. Эвристическое, статистическое, энтропийное определение информативности.
18. Асимптотическая эквивалентность статистического и энтропийного определения. Сравнение областей эвристических и статистических закономерностей.
19. Разновидности закономерностей: конъюнкции пороговых предикатов (гиперпараллелепипеды), синдромные правила, шары, гиперплоскости.
20. Градиентный алгоритм синтеза конъюнкций, частные случаи: жадный алгоритм, стохастический локальный поиск, стабилизация, редукция. Бинаризация признаков.
21. Решающие деревья для задач классификации и регрессии.
22. Линейный классификатор, непрерывные аппроксимации пороговой функции потерь. Связь с методом максимума правдоподобия.
23. Метод стохастического градиента и частные случаи: адаптивный линейный элемент ADALINE, персептрон Розенблатта, правило Хэбба.
24. Теорема Новикова о сходимости. Доказательство теоремы Новикова.
25. Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, "выбивание" из локальных минимумов.
26. Метод стохастического среднего градиента SAG.
27. Проблема мультиколлинеарности и переобучения, редукция весов (weight decay).
28. Байесовская регуляризация. Принцип максимума совместного правдоподобия данных и модели. Квадратичный (гауссовский) и лапласовский регуляризаторы.
29. Настройка порога решающего правила по критерию числа ошибок I и II рода.
30. Кривая ошибок (ROC curve). Алгоритм эффективного построения ROC-кривой. Градиентный метод максимизации AUC.
31. . Понятие опорных векторов. Рекомендации по выбору константы C. Функция ядра (kernel functions), спрямляющее пространство, теорема Мерсера.
32. Способы конструктивного построения ядер. Примеры ядер. Обучение SVM методом активных ограничений.

33. SVM - регрессия. Метод релевантных векторов RVM. Регуляризации для отбора признаков: LASSO SVM, Elastic Net SVM, SFM, RFM.
34. Метод наименьших квадратов, его вероятностный смысл и геометрический смысл.
35. Сингулярное разложение. Проблемы мультиколлинеарности и переобучения. Регуляризация.
36. . Гребневая регрессия. Лассо Тибширани, сравнение с гребневой регрессией.
37. Метод главных компонент и декоррелирующее преобразование Карунена - Лоэва, его связь с сингулярным разложением.
38. Линейные композиции, бустинг Основные понятия: базовый алгоритм (алгоритмический оператор), корректирующая операция. Взвешенное голосование.
39. Алгоритм AdaBoost. Процесс последовательного обучения базовых алгоритмов.
40. Теорема о сходимости бустинга. Базовые алгоритмы в бустинге. Решающие пни. Градиентный бустинг.
41. Стохастические методы: бэггинг и метод случайных подпространств. Случайные леса.
42. Оптимальный байесовский классификатор. Принцип максимума апостериорной вероятности. Функционал среднего риска.
43. Ошибки I и II рода. Теорема об оптимальности байесовского классификатора.
44. Оценивание плотности распределения: три основных подхода.
45. Наивный байесовский классификатор.

15. ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. № АК-44/05вн) в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации обучающихся с ОВЗ с целью оказания помощи в установлении полноценных межличностных отношений.

Обучение лиц с ограниченными возможностями здоровья осуществляется с учетом индивидуальных психофизических особенностей, а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида.

Для лиц с нарушением слуха возможно предоставление информации визуально (краткий конспект лекций, основная и дополнительная литература), на лекционных и практических занятиях допускается присутствие ассистента, а так же, сурдопереводчиков и тифлосурдопереводчиков.

Оценка знаний обучающихся на практических занятиях осуществляется на основе письменных конспектов ответов на вопросы, письменно выполненных практических заданий.

Доклад так же может быть предоставлен в письменной форме (в виде реферата), при этом требования к содержанию остаются теми же, а требования к качеству изложения материала (понятность, качество речи, взаимодействие с аудиторией и т. д.) заменяются на соответствующие требования, предъявляемые к письменным работам (качество оформления текста и списка литературы, грамотность, наличие иллюстрационных материалов и т.д.)

С учетом состояния здоровья просмотр кинофильма с последующим анализом может быть проведен дома (например, при необходимости дополнительной звукоусиливающей аппаратуры (наушники)). В таком случае обучающийся предоставляет письменный анализ, соответствующий предъявляемым требованиям.

Промежуточная аттестация для лиц с нарушениями слуха проводится в письменной форме, при этом используются общие критерии оценивания. При необходимости, время подготовки на зачете может быть увеличено.

Для лиц с нарушением зрения допускается аудиальное предоставление информации (например, с использованием программ-синтезаторов речи), а так же использование на лекциях звукозаписывающих устройств (диктофонов и т.д.). Допускается присутствие на занятиях ассистента (помощника), оказывающего обучающимся необходимую техническую помощь.

Оценка знаний обучающихся на практических занятиях осуществляется в устной форме (как ответы на вопросы, так и практические задания). При необходимости анализа фильма может быть заменен описанием ситуации межэтнического взаимодействия (на основе опыта респондента, художественной литературы и т.д.), позволяющим оценить степень сформированности навыков владения методами анализа и выявления специфики функционирования и развития психики, позволяющими учитывать влияние этнических факторов. При проведении промежуточной аттестации для лиц с нарушением зрения тестирование может быть заменено на устное собеседование по вопросам.

Лица с нарушениями опорно-двигательного аппарата не нуждаются в особых формах предоставления учебных материалов. Однако, с учетом состояния здоровья часть занятий может быть реализована дистанционно (при помощи сети «Интернет»). Так, при невозможности посещения лекционного занятия обучающийся может воспользоваться кратким конспектом лекции.

При невозможности посещения практического занятия обучающийся должен предоставить письменный конспект ответов на вопросы, письменно выполненное практическое задание.

Доклад так же может быть предоставлен в письменной форме (в виде реферата), при этом

требования к содержанию остаются теми же, а требования к качеству изложения материала (понятность, качество речи, взаимодействие с аудиторией и т. д.) заменяются на соответствующие требования, предъявляемые к письменным работам (качество оформления текста и списка литературы, грамотность, наличие иллюстрационных материалов и т.д.).

Промежуточная аттестация для лиц с нарушениями опорно-двигательного аппарата проводится на общих основаниях, при необходимости процедура зачета может быть реализована дистанционно (например, при помощи программы Skype).

Для этого по договоренности с преподавателем обучающийся в определенное время выходит на связь для проведения процедуры зачета. В таком случае зачет сдается в виде собеседования по вопросам (см. формы проведения промежуточной аттестации для лиц с нарушениями зрения). Вопрос и практическое задание выбираются самим преподавателем.

Примечание: Фонды оценочных средств, включающие типовые задания и методы оценки, критерии оценивания, позволяющие оценить результаты освоения данной дисциплины обучающимися с ОВЗ могут входить в состав РПД на правах отдельного документа.

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