

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»



УТВЕРЖДАЮ

Директор по образовательной деятельности

С.Т. Князев

С.Т. Князев

2021 г.

РАБОЧАЯ ПРОГРАММА МОДУЛЯ

Код модуля
М.1.13

Модуль
Спортивный анализ данных

Екатеринбург, 2021

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Инженерия искусственного интеллекта	Код ОП 09.04.01
Направление подготовки Информатика и вычислительная техника	Код направления и уровня подготовки 09.04.01

Области образования, в рамках которых реализуется модуль образовательной программы по СУОС УрФУ:

№ п/п	Перечень областей образования, для которых разработан СУОС УрФУ	Уровень подготовки
1.	Инженерное дело, технологии и технические науки	магистратура

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

Рекомендовано учебно-методическим советом института радиозлектроники и информационных технологий - РТФ

Протокол № 7 от 11 октября 2021 г.

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Спортивный анализ данных

1.1. Аннотация содержания модуля

Модуль «Спортивный анализ данных» состоит из одноименной дисциплины.

Цель освоения дисциплины «Спортивный анализ данных» ознакомить студентов с современной платформой для проведения соревнований "Kaggle .com". Студенты узнают, как использовать различные алгоритмы и методы анализа данных для решения конкретных прикладных задач. На практике рассматриваются все типы задач анализа данных: анализ табличных данных, анализ временных рядов, обработка естественного языка, обработка изображений. В курсе рассматриваются методы анализа базовых обученных моделей, выбора и проверки новых возможностей, оптимальные методы поиска лучшего алгоритма решения задачи.

1.2. Структура и объем модуля

Таблица 1

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах
1.	Спортивный анализ данных	6 з.е. / 216 ч.
ИТОГО по модулю:		6 з.е. / 216 ч.

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	<i>отсутствуют</i>
Постреквизиты и корреквизиты модуля	<i>отсутствуют</i>

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
1	2	3	4
Спортивный анализ данных	ПК-6. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях	ПК-6.1. Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях	ПК-6.1. 3-1. Знает методологию и принципы руководства проектом по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных

			<p>ПК-6.1. З-2. Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных</p> <p>ПК-6.1. У-1. Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных</p> <p>ПК-6.1. У-2. Умеет сосредотачивать внимание на целях, достижение которых обеспечивает большую отдачу и сильное воздействие</p> <p>ПК-6.1. У-3. Умеет формировать матрицу приоритетов, включая критерии отбора проектов для реализации</p>
--	--	--	--

1.5. Форма обучения

Обучение по дисциплине модуля может осуществляться в очной форме.

2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН МОДУЛЯ

**ПРОГРАММА МОДУЛЯ
СПОРТИВНЫЙ АНАЛИЗ ДАННЫХ**

**РАЗДЕЛ 2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН
МОДУЛЯ**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
СПОРТИВНЫЙ АНАЛИЗ ДАННЫХ**

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ- РТФ, УрФУ

Рекомендовано учебно-методическим советом института радиоэлектроники и информационных технологий - РТФ

Протокол № 7 от 11 октября 2021 г.

2. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ Спортивный анализ данных

2.1. Технологии обучения, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Смешанная модель обучения с использованием онлайн-курса УрФУ;

2.2. Содержание дисциплины

Таблица 1.3

Код раздела, темы	Раздел, тема дисциплины*	Содержание
1	Обзор инструментов применяемых для анализа данных и машинного обучения	Знакомство с инструментами: jupyter notebook, google colaboratory, kaggle kernel. Базовый функционал и приемы работы. Основные понятия Data Science и Machine Learning
2	Обзор библиотеки Numpy	Применение библиотеки numpy для операции над тензорами и решения задач линейной алгебры и генерации тензоров различной размерности по заданным правилам
3	Работа с библиотекой Pandas (Matplotlib, seaborn)	Библиотека для работы с табличными данными pandas: чтение и запись файлов, методы просмотра данных, индексация, срезы, фильтрация, агрегация и сортировка данных, визуализация данных, разведывательный анализ данных, базовые статистики.
4	Модели машинного обучения для решения задач классификации	Постановка задачи классификации. Изучение алгоритмов классификации в машинном обучении: логистическая регрессия, машины опорных векторов, наивный Байес, K ближайших соседей, деревья решений, ансамблевые модели
5	Модели машинного обучения для решения задач регрессии	Постановка задачи регрессии. Изучение алгоритмов регрессии в машинном обучении: линейная регрессия, машины опорных векторов, K ближайших соседей, деревья решений, ансамблевые модели. Регуляризация моделей машинного обучения
6	Метрики качества при решении задач классификации и регрессии	Изучение метрик качества моделей машинного обучения для классификации и регрессии. Рассмотрение сильных и слабых сторон различных метрик для различных задач. Применение изученных метрик
7	Предварительная обработка данных при решении задач анализа данных	Методы предварительной обработки: заполнение пропусков в данных, обработка выбросов в данных.
8	Преобразование и создание новых признаков для решения задач анализа данных	Изменение непрерывных типов данных: логарифмирование, построение полиномов, дискретизация данных, стандартизация и нормализация данных. Различные методы преобразование категориальных данных.
9	Методы выбора лучших признаков для решения задач классификации и регрессии	Методы отбора лучших для решения поставленной задачи признаков в задачах классификации и регрессии: на основе статистических подходов и методов машинного обучения
10	Методы выбора лучших моделей и алгоритмов для	Приемы отбора лучших моделей для решения поставленных задач. Различные методики разбиения

	решения поставленных задач анализа данных	выборки для более стабильной и правдоподобной оценки
11	Методы подбора гиперпараметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных	Инструменты построения конвейеров обработки данных в машинном обучении. Методы и инструменты автоматического подбора гиперпараметров моделей машинного обучения, а также подбора самих моделей и методов предварительной обработки
12	Работа с временными рядами	Понятие временного ряда и его составляющие элементы. Методы прогнозирования на основе эконометрических подходов. Методы прогнозирования на основе регрессионного подхода и моделей машинного обучения. Генерация новых временных признаков для временного ряда.
13	Задачи обучения без учителя: понижение размерности	Обзор задач обучения без учителя и их применимость на практике. Методы понижения размерности и их практические приложения
14	Задачи обучения без учителя: кластеризация	Обзор методов кластеризации и их сравнительные характеристики. Методы оценки оптимального числа кластеров. Кластеризация для сегментации и анализа. Кластеризация, как метод сжатия информации и снижения размерности.
15	Задачи обучения без учителя: поиск аномалий	Постановка задачи поиска аномалий и выявления новизны. Методы поиска аномалий на основе подходов машинного обучения.
16	Обработка естественного языка: предварительная обработка	Знакомство с задачами обработки естественного языка.: Методы первичной предварительной обработки текстов: очистка, токенизация, лемматизация, стемминг.
17	Обработка естественного языка: векторные модели	Преобразование текста в векторное пространство. Статистические методы построения векторных моделей: мешок слов, ONE, tf-idf. Тематическое моделирование: латентное размещение Дирихле (LDA), латентный семантический анализ (LSA). Нейросетевые модели: word2vec, fastText, GloVe.
18	Обработка естественного языка: типы решаемых задач	Рассмотрение спектра задач обработки естественного языка: моделирование языка, классификация, поиск именованных сущностей, суммаризация, генерация.
19	Введение в нейронные сети	Базовые понятия нейронных сетей: нейрон, обучение нейрона, нейронная сеть, обучение нейронной сети. Функции потерь и метрики качества.
20	Нейронные сети для решения задач регрессии	Применение нейронных сетей для решения задач регрессии. Специфические для регрессии функции активации нейронов и метрики качества. Методы предварительной обработки данных.
21	Нейронные сети для решения задач классификации	Применение нейронных сетей для решения задач классификации. Специфические для классификации функции активации нейронов и метрики качества. Методы предварительной обработки данных.
22	Нейронные сети для решения задач обработки изображений	Знакомство с задачами обработки изображений. Изучение сверточных нейронных сетей. Понятие ядра, свертки и подвыборки.
23	Нейронные сети для решения задач обработки изображений: современные	Изучение современных архитектур нейронных сетей для решения задач обработки изображений: VGG-16, VGG-19, ResNet, Inception, Xception, DenseNet,

	архитектуры	MobileNet, EfficientNet.
24	Нейронные сети для решения задач обработки естественного языка	Архитектуры нейронных сетей для решения задач обработки естественного языка: одномерные сверточные сети, рекуррентные нейронные сети.
25	Нейронные сети для решения задач обработки естественного языка: современные архитектуры	Обзор современных модификаций нейронных сетей для решения задач обработки естественного языка: трансформеры, механизм внимания, BERT
26	Нейронные сети для решения задач предсказания временных рядов	Применение нейронных сетей для решения задачи прогнозирования временных рядов: одномерные сверточные нейронные сети, рекуррентные нейронные сети. WaveNet для прогнозирования временного ряда

2.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

2.4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ Спортивный анализ данных

Электронные ресурсы (издания)

1. Хливненко, Л. В. Практика нейросетевого моделирования : учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 200 с. — ISBN 978-5-8114-8264-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/173811> (дата обращения: 05.10.2021).
2. Клетте, Р. Компьютерное зрение. Теория и алгоритмы : учебник / Р. Клетте ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 506 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131691> (дата обращения: 08.10.2021).
3. Тарасов, И. Е. Статистический анализ данных в информационных системах : учебно-методическое пособие / И. Е. Тарасов. — Москва : РТУ МИРЭА, 2020. — 96 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/163854> (дата обращения: 06.10.2021).
4. Ростовцев, В. С. Искусственные нейронные сети : учебник / В. С. Ростовцев. — Санкт-Петербург : Лань, 2019. — 216 с. — ISBN 978-5-8114-3768-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/122180> (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.

Печатные издания

1. Жерон, Орельен, Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем, 2-е изд. Пер. с англ. - СПб.: ООО "Диалектика": 2020. - 1040 с.: ил. - Парал. тит. англ.
2. Копец Дэвид, Классические задачи Computer Science на языке Python. - СПб.: Питер, 2020. - 256 с.: ил. - (Серия «Библиотека программиста»).
3. Элбон Крис, Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. — СПб.: БХВ-Петербург, 2019. — 384 с.: ил.
4. Вейдман Сет, Глубокое обучение: легкая разработка проектов на Python. — СПб.: Питер, 2021. — 272 с.: ил. — (Серия «Бестселлеры O'Reilly»).
5. Микелуччи У., Прикладное глубокое обучение. Подход к пониманию глубоких нейронных сетей на основе метода кейсов: Пер. с англ. - СПб.: БХВ-Петербург, 2020. - 368 с.: ил.
6. Шолле Франсуа, Глубокое обучение на Python. — СПб.: Питер, 2018. — 400 с.: ил. — (Серия «Библиотека программиста»).

7. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.
8. Хобсон Лейн, Ханнес Хапке, Коул Ховард, Обработка естественного языка в действии. — СПб.: Питер, 2020. — 576 с.: ил. — (Серия «Для профессионалов»)
9. Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони, Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. — СПб.: Питер, 2019. — 368 с.: ил. — (Серия «Бестселлеры O'Reilly»).

Профессиональные базы данных, информационно-справочные системы

1. Applied Science & Technology Source. EBSCO publishing <http://search.ebscohost.com>
2. Wiley Online Library <http://onlinelibrary.wiley.com/>
3. Гугл Академия <https://scholar.google.ru/>

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

1. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
2. eBook Collections Springer Nature – <https://link.springer.com/>
3. Гугл Академия – <https://scholar.google.ru/>
4. Электронный научный архив УрФУ <https://elar.urfu.ru/>
5. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
6. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
7. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
8. Университетская библиотека ONLINE – <https://biblioclub.ru/>
9. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
10. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
11. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

2.5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Спортивный анализ данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
1	Лекции; Практические занятия.	Компьютерный класс. Мультимедийный проектор с экраном.; Сетевое оборудование.;	MS PowerPoint, MS Word, MS Excel. MS Teams Используется бесплатно-

		Локальная сеть с выходом в глобальную сеть Internet.	распространяемое программное обеспечение: 1. Python – https://www.python.org/
--	--	--	--

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Код модуля
М.1.13

Модуль
Спортивный анализ данных

Екатеринбург, 2021

Оценочные материалы по модулю составлены авторами:

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. СТРУКТУРА И ОБЪЕМ МОДУЛЯ Спортивный анализ данных

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах	Форма итоговой промежуточной аттестации по дисциплинам модуля и в целом по модулю
1.	Спортивный анализ данных	6 / 216	<i>Зачет (2-3 семестр)</i>
ИТОГО по модулю:		6 / 216	

2. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО МОДУЛЮ

Не предусмотрено.

Раздел 3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ 1
СПОРТИВНЫЙ АНАЛИЗ ДАННЫХ

Модуль М.1.13 Спортивный анализ данных

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Спортивный анализ данных

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Индикаторы должны учитываться при выборе и составлении заданий контрольно-оценочных мероприятий (оценочных средств) текущей и промежуточной аттестации.

Таблица 1.1

Код и наименование компетенций, формируемые с участием дисциплины	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3	4
ПК-6. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях	ПК-6.1. Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях	<p>ПК-6.1. 3-1. Знает методологию и принципы руководства проектом по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных</p> <p>ПК-6.1. 3-2. Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных</p> <p>ПК-6.1. У-1. Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных</p> <p>ПК-6.1. У-2. Умеет сосредотачивать внимание на целях, достижение которых обеспечивает</p>	<p>1) Контрольные работы (2 шт.)</p> <p>2) Домашние работы (4 шт.)</p> <p>3) Выполнение практических работ</p> <p>4) Зачет (2 семестр, 3 семестр).</p>

		большую отдачу и сильное воздействие	
		ПК-6.1. У-3. Умеет формировать матрицу приоритетов, включая критерии отбора проектов для реализации	

2. ВИДЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ, ВКЛЮЧАЯ МЕРОПРИЯТИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ

2.1. Распределение объема времени по видам учебной работы

Таблица 2

№ п/ п	Наименование дисциплины модуля Спортивный анализ данных	Объем времени, отведенный на освоение дисциплины модуля								
		Аудиторные занятия, час.				Промежуточная аттестация (форма итогового контроля)	Контактная работа (час.)	Самостоятельная работа студента, включая текущую аттестацию (час.)	Всего по дисциплине	
		Занятия лекцион ного типа	Практиче ские работы	Лаборатор ные работы	Всего				Час.	Зач. ед.
1	2	3	4	5	6	7	8	9	10	11
1.	Спортивный анализ данных	36	36	0	72	Зачет (2-3 семестр)	83.3	132.7	216	6

2.2. Виды СРС, количество и объем времени на контрольно-оценочные мероприятия СРС по дисциплине

Контрольно-оценочные мероприятия СРС включают самостоятельное изучение материала, подготовку к аудиторным занятиям и мероприятиям текущего контроля, выполнение и оформление внеаудиторных мероприятий текущего контроля и подготовку к мероприятиям промежуточного контроля.

Таблица 3

№ п/п	Вид самостоятельной работы студента по дисциплине модуля	Количество контрольно-оценочных мероприятий СРС	Объем контрольно-оценочных мероприятий СРС (час.)
1.	Подготовка к аудиторным занятиям и мероприятиям текущего контроля: лекционным, практическим занятиям.		27 час.
2.	Выполнение и оформление мероприятий текущего контроля:		
2.1	Контрольная работа	2	10 час.
2.2	Домашняя работа №1	1	5 час.
2.3	Домашняя работа №2	1	5 час.
2.4	Домашняя работа №3	1	5 час.
2.5	Домашняя работа №4	1	5 час.
3.	Подготовка к зачету	зачет (2 и 3 семестр)	24 час.
4.	Самостоятельное изучение материала		51.7 час.
Итого на СРС по дисциплине:			132.7 час.

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

Процедуры текущей и промежуточной аттестации по дисциплине

2 семестр

1.Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5

Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Контрольная работа</i>	2 сем.	70
<i>Самостоятельное изучение материала</i>	2 сем.	30
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – <i>Зачет</i>		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение и оформление практических работ</i>	2 сем.	50
<i>Домашняя работа №1</i>	2 сем.	25
<i>Домашняя работа №2</i>	2 сем.	25
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям– не предусмотрена		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0		
3. Лабораторные занятия: Не предусмотрены		
коэффициент значимости совокупных результатов лабораторных занятий – 0		

3 семестр

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Контрольная работа</i>	3 сем.	70
<i>Самостоятельное изучение материала</i>	3 сем.	30
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – <i>Зачет</i>		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение и оформление практических работ</i>	3 сем.	50
<i>Домашняя работа</i>	3 сем.	25
<i>Домашняя работа</i>	3 сем.	25
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям– не предусмотрена		

Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0

**3. Лабораторные занятия: Не предусмотрены
коэффициент значимости совокупных результатов лабораторных занятий – 0**

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Личностные качества	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2. Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
1.	Результаты обучения	Отлично	Зачтено	Высокий (В)

	(индикаторы) достигнуты в полном объеме, замечаний нет	(80-100 баллов)		
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

Задания по контрольно-оценочным мероприятиям в рамках текущей и промежуточной аттестации должны обеспечивать освоение и достижение результатов обучения (индикаторов) и предметного содержания дисциплины на соответствующем уровне.

5.1. Описание контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

5.1.1. Практические занятия

	Номер занятия	Примерный перечень тем практических занятий
2 семестр	1	Обзор инструментов применяемых для анализа данных и машинного обучения. Обзор библиотеки Numpy. Работа с библиотекой Pandas (Matplotlib, seaborn)
	2	Модели машинного обучения для решения задач классификации
	3	Модели машинного обучения для решения задач регрессии
	4	Метрики качества при решении задач классификации и регрессии
	5	Предварительная обработка данных при решении задач анализа данных. Преобразование и создание новых признаков для решения задач анализа данных
	6	Методы выбора лучших признаков для решения задач классификации и регрессии
	7	Методы подбора гиперпараметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных
	8	Работа с временными рядами

	9	Задачи обучения без учителя: понижение размерности Задачи обучения без учителя: поиск аномалий
3 семестр	1	Обработка естественного языка: предварительная обработка
	2	Обработка естественного языка: векторные модели
	3	Введение в нейронные сети. Нейронные сети для решения задач регрессии
	4	Нейронные сети для решения задач классификации
	5	Нейронные сети для решения задач обработки изображений
	6	Нейронные сети для решения задач обработки изображений: современные архитектуры
	7	Нейронные сети для решения задач обработки естественного языка
	8	Нейронные сети для решения задач обработки естественного языка: современные архитектуры
	9	Нейронные сети для решения задач предсказания временных рядов

5.1.2. Лабораторные занятия

Не предусмотрено.

5.1.3. Курсовая работа / Курсовой проект

Не предусмотрено.

5.1.4. Контрольная работа

Примерная тематика контрольных работ:

Контрольная работа №1

Присоединитесь к соревнованиям на платформе Kaggle.com, проведите первичный анализ данных соревнования, выполните предварительную обработку данных и применяя различные модели машинного обучения добейтесь качества решения указанного в условии задания. Решается задача классификации объектов.

Контрольная работа №2

Присоединитесь к соревнованиям на платформе Kaggle.com, проведите первичный анализ данных соревнования, выполните предварительную обработку данных и применяя различные модели машинного обучения добейтесь качества решения указанного в условии задания. Решается задача регрессии.

Примерные задания в составе контрольных работ:

Перейдите по ссылке <https://www.kaggle.com/t/993ac81ad69a477580bd8cf1beca9a32>

Там вы увидите соревнование на платформе kaggle.com

На вкладке «Overview» вы увидите описание задания.

На вкладке «Data» находятся все необходимые данные, которые будут использованы в соревновании.

На вкладке «Notebooks» в разделе «Public» находится базовое решение с примером кода «BaseLine_1_5».

Скопировав базовое решение и запустив его у себя, вы сможете достигнуть точности около значения 0,83.

Ваша задача изменяя любые параметры архитектуры нейронной сети или параметры обучения

нейронной сети достигнуть наибольшего качества модели.

Результат своих экспериментов вы сможете наблюдать на вкладке «Leaderboard» сразу после загрузки ответа во вкладке «My Submissions».

Ваша цель достигнуть качества модели на тестовых данных 0.9 и выше. Тогда вы проходите задание

Когда, используя LeaderBoard убедитесь, что результат вас устраивает, скачайте файл своего решения, который вы можете увидеть, открыв свой рабочий блокнот в режиме просмотра (но не редактирования) и перейдя справа во вкладку «Output».

Скачанный файл в формате «.csv» загрузите сюда, на платформу курса.

5.1.5. Домашняя работа

Примерная тематика домашних работ:

Домашняя работа №1:

Дан colab notebook с заданиями по библиотеке numpy. Выполните все задания и отправьте ссылку на свое решение в google forms

Примерные тексты заданий:

1. Создайте одномерный массив numpy из 10 элементов заполненный числом 1.5
2. Создайте двумерный массив numpy из 25 элементов заполненный числом 0
3. Создайте вектор длиной 15 со случайными целыми числами в диапазоне от -10 до 10. Сделайте реверс вектора
4. Создайте матрицу 5 на 5 со случайными целыми числами в диапазоне от -15 до 15. Замените отрицательные элементы на их квадраты
5. Даны две матрицы. Убедитесь, что их можно перемножить. Произведите операцию умножения матриц, если это возможно, иначе выведите ошибку
6. Дан вектор, проверить есть в нем Nan. Проверить есть в нем знак бесконечности. Заполнить их нулями

Домашняя работа №2:

Дан colab notebook с заданиями по библиотеке pandas. Выполните все задания и отправьте ссылку на свое решение в google forms

Примерные тексты заданий:

Дан набор данных с демографической информацией о людях, получающих больше 50 тысяч долларов в год и меньше. Проанализируйте набор и ответьте на вопросы используя библиотеку pandas.

1. Сколько мужчин и женщин (признак sex) представлено в этом наборе данных?
2. Каков средний возраст (признак age) женщин?
3. Какова доля граждан Германии (признак native-country)?
4. Правда ли, что люди, которые получают больше 50k, имеют как минимум высшее образование? (признак education – Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters или Doctorate)
5. Среди кого больше доля зарабатывающих много (>50K): среди женатых или холостых мужчин (признак marital-status)? Женатыми считаем тех, у кого marital-status начинается с Married (Married-civ-spouse, Married-spouse-absent или Married-AF-spouse), остальных считаем холостыми.

6. Какое максимальное число часов человек работает в неделю (признак hours-per-week)? Сколько людей работают такое количество часов и каков среди них процент зарабатывающих много?

Домашняя работа №3:

Дан colaboratory notebook с заданиями по библиотекам pandas, matplotlib, seaborn. Выполните все задания и отправьте ссылку на свое решение в google forms.

Примерные тексты заданий:

Дан набор данных с информацией об олимпийских играх за разные годы. Проведите первичный визуальный анализ.

1. Постройте графики: линейные, столбчатые, круговые, гистограммы, ящики с усами – по каждому столбцу данных
2. Постройте точечные графики, отражающие взаимосвязь между различными признаками
3. Проведите корреляционный анализ и визуализируйте его с помощью тепловой карты
4. Сделайте краткое описание к каждому графику, которое отразит информацию, которую вы увидели на визуализации.

Домашняя работа №4:

Дан colaboratory notebook с заданиями по библиотекам sklearn. Выполните все задания и отправьте ссылку на свое решение в google forms

Примерные тексты заданий:

Дан набор данных с информацией о стоимостях квартир в зависимости от их параметров. Выполните следующие задания.

1. Проведите первичный анализ: статистический и визуальный и презентуйте его.
2. Проведите предварительную обработку данных: заполнение пропусков, обработка выбросов, генерация новых признаков.
3. Примените различные модели машинного обучения и добейтесь указанного в задании качества модели по заданной метрике.

5.1.6. Расчетная работа / Расчетно-графическая работа

Не предусмотрено.

5.1.7. Реферат / эссе / творческая работа

Не предусмотрено.

5.1.8. Проектная работа

Не предусмотрено.

5.1.9. Деловая (ролевая) игра / Дебаты / Дискуссия / Круглый стол

Не предусмотрено.

5.1.10. Кейс-анализ

Не предусмотрено.

5.2. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.2.1. Зачет в форме независимого тестового контроля (НТК)

НТК по дисциплине модуля не проводится.

5.2.2. Зачет в традиционной форме (устные /письменные ответы на вопросы)

Список примерных вопросов для Зачета (2 семестр):

1. Назовите ключевые типы данных, которые могут быть в задачах интеллектуального анализа данных.
2. Назовите задачи, которые могут решаться при обучении моделей машинного обучения с учителем. Приведите примеры таких задач.
3. Назовите основные этапы решения задачи интеллектуального анализа данных.
4. Опишите, какую предварительную обработку и подготовку данных необходимо провести при работе с задачами на изображениях.
5. Какие инструменты и модели подходят для работы с изображениями?
6. Опишите, какую предварительную обработку и подготовку данных необходимо провести при работе с задачами на текстовых данных.
7. Какие инструменты и модели подходят для работы с текстовыми данными?
8. Опишите, какую предварительную обработку и подготовку данных необходимо провести при работе с задачами на табличных данных.
9. Какие инструменты и модели подходят для работы с табличными данными?
10. Что из себя представляют временные ряды и в чем особенность работы с задачами на предсказание временных рядов?
11. Какие модели машинного обучения применяются для решения задач на предсказание временных рядов?
12. Что такое векторное представление текста? Зачем оно нужно? Какие векторные модели вы знаете?
13. Назовите подходы к обработке пропусков в данных. Какие методики заполнения пропусков вы знаете? Какие у них особенности
14. Что такое нормализация и шкалирование данных? Для чего они нужны в задачах анализа данных?
15. Какие подходы к кодированию категориальных признаков вы знаете? Приведите примеры, когда приемлемо использовать каждый из подходов.
16. Какие признаки для обучения модели машинного обучения можно извлечь из типов данных «timestamp» и «datetime»?
17. Какие признаки можно извлечь из временного ряда, чтобы получить возможность решить задачу временного ряда посредством линейной модели?
18. Назовите методы увеличения обучающей выборки в случае решения задачи анализа данных на изображениях.
19. Опишите ключевые метрики качества моделей машинного обучения при решении задачи классификации. Какие из них не чувствительны к несбалансированным данным?
20. Опишите ключевые метрики качества моделей машинного обучения при решении задачи регрессии
21. Опишите процесс выбора признаков, которые положительно влияют на качество итоговой модели машинного обучения.
22. Для чего нужна оценка качества модели. Какие подходы к проверке качества моделей вы знаете? Назовите преимущества и недостатки основных подходов.
23. Назовите типы ансамблевых моделей машинного обучения.
24. Расскажите особенности работы с фреймворком XGBoost
25. Расскажите особенности работы с фреймворком LightGBM
26. Расскажите особенности работы с фреймворком CatBoost
27. Опишите устройство временного ряда.
28. Какие инструменты для прогнозирования временного ряда вы знаете?

29. Опишите процесс предварительной обработки текстовой информации.
30. Расскажите, что такое tf-idf.

Список примерных вопросов для Зачета (3 семестр):

1. Опишите устройство искусственного нейрона
2. Опишите процесс обучения искусственного нейрона
3. Опишите функции активации, которые вы знаете. Проведите их сравнительный анализ
4. Опишите устройство полносвязной нейронной сети.
5. Какие задачи может решать нейронная сеть? Как изменяется выходной слой и функция активации на выходном слое в зависимости от решаемой задачи?
6. Опишите процесс обучения нейронной сети. На чем основан метод обратного распространения ошибки?
7. Расскажите про метрики качества нейронных сетей для решения различных задач. Проведите сравнительный анализ
8. Расскажите про функции потерь нейронных сетей для решения различных задач. Проведите сравнительный анализ
9. Что такое переобучение нейронной сети? Какие методы борьбы с переобучением вы знаете? Как можно выявить факт переобучения?
10. Какие гиперпараметры нейронной сети вы знаете? На что они влияют? Как происходит процесс подбора гиперпараметров?
11. Что такое градиентный спуск? Опишите как работает градиентный спуск и стохастический градиентный спуск.
12. Опишите необходимые свойства функций ошибок, для которых можно применить градиентный спуск.
13. Какие модификации градиентного спуска вы знаете? Проведите сравнительный анализ модификаций градиентного спуска
14. Опишите недостатки полносвязных нейронных сетей для решения задач обработки изображений.
15. Опишите устройство сверточной нейронной сети.
16. Опишите принцип работы сверточного слоя.
17. Для чего нужны ядра в сверточном слое? Как они подбираются?
18. Как происходит процесс переноса обучения в сверточных нейронных сетях?
19. Какие современные архитектуры нейронных сетей для решения задач обработки изображений вы знаете? Проведите краткий сравнительный анализ.
20. Опишите принцип работы нейронной сети с архитектурой ResNet
21. Опишите принцип работы нейронной сети с архитектурой Inception
22. Опишите принцип работы нейронной сети с архитектурой Xception
23. Опишите принцип работы нейронной сети с архитектурой VGG-16/19
24. Опишите принцип работы рекуррентных нейронных сетей? Какие задачи они решают?
25. Опишите принцип работы LSTM блока в нейронной сети.
26. Опишите принцип работы GRU блока в нейронной сети.
27. Опишите принцип работы двунаправленных рекуррентных нейронных сетей. Для чего они применяются?
28. Какие задачи решают нейронные сети при работе с изображениями?
29. Какие задачи решают нейронные сети при работе с текстовой информацией?
30. Какие векторные модели текстов вы знаете? Проведите их сравнительный анализ
31. Что такое word2vec? Опишите процесс его обучения и особенности применения.