

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)

Институт информационных технологий и радиоэлектроники



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Спортивный анализ данных»

направление подготовки / специальность
09.04.04 «Программная инженерия»

направленность (профиль) подготовки
Инженерия искусственного интеллекта

г. Владимир
2021

1. ЦЕЛИ ОСВОЕНИЯ 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.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина Спортивный анализ данных относится к части учебного плана, формируемой участниками образовательных отношений.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОПОП (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине, в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	
ПК-7. Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях	ПК-7.1. Знать: ПК-7.1.1. принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» ПК-7.1.2. принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	Знает: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»; современное состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта Умеет: руководить проектами по созданию, внедрению и поддержке систем искусственного	1) Контрольные работы 2) Рейтинг-контроль 3) Выполнение лабораторных работ 4) Зачет с оценкой

	<p>ПК-7.1.3. современное состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта</p> <p>ПК-7.2. Уметь:</p> <p>ПК-7.2.1. руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение»</p> <p>ПК-7.2.2. руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»</p> <p>ПК-7.3. Иметь навыки:</p> <p>ПК-7.3.1. проведения анализа новых направлений, методов и технологий в области искусственного интеллекта и определения наиболее перспективных для различных областей применения</p>	<p>интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»</p> <p>Имеет навыки:</p> <p>проведения анализа новых направлений, методов и технологий в области искусственного интеллекта и определения наиболее перспективных для различных областей применения</p>	
--	---	---	--

4. ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 4 зачетных единицы, 144 часа

Тематический план форма обучения – очная

№ п/п	Наименование тем и/или разделов/тем дисциплины	Семестр	Неделя семестра	Контактная работа обучающихся с педагогическим работником				Самостоятельная работа	Формы текущего контроля успеваемости, форма промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	Лабораторные работы	в форме практической подготовки		
1	Обзор инструментов, применяемых для анализа данных и машинного обучения. Обзор библиотеки Numpy Работа с библиотекой Pandas (Matplotlib, seaborn)	3	1-2	2		2	1	12	
2	Модели машинного обучения для решения задач классификации Модели машинного обучения для решения задач регрессии Метрики качества при решении задач классификации и регрессии Предварительная обработка данных при решении задач анализа данных Преобразование и создание новых признаков для решения задач анализа данных	3	3-4	2		2	1	12	
3	Методы выбора лучших признаков для решения задач классификации и регрессии Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных Методы подбора гипер параметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных	3	5-6	2		2	1	12	Рейтинг-контроль №1
4	Работа с временными рядами	3	7-8	2		2	1	12	

5	Задачи обучения без учителя: понижение размерности Задачи обучения без учителя: кластеризация Задачи обучения без учителя: поиск аномалий	3	9-10	2		2	1	12	
6	Обработка естественного языка: предварительная обработка Обработка естественного языка: векторные модели Обработка естественного языка: типы решаемых задач	3	11-12	2		2	1	12	Рейтинг-контроль №2
7	Введение в нейронные сети Нейронные сети для решения задач регрессии Нейронные сети для решения задач классификации Нейронные сети для решения задач обработки изображений Нейронные сети для решения задач обработки изображений: современные архитектуры	3	13-14	2		2	1	12	
8	Нейронные сети для решения задач обработки естественного языка Нейронные сети для решения задач обработки естественного языка: современные архитектуры	3	15-16	2		2	1	12	
9	Нейронные сети для решения задач предсказания временных рядов	3	17-18	2		2	1	12	Рейтинг-контроль №3
Всего за 3 семестр:				18		18		108	Зачет с оценкой
Наличие в дисциплине КП/КР									
Итого по дисциплине				18		18		108	Зачет с оценкой

Содержание лекционных занятий по дисциплине

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

Обзор библиотеки NumPy

Применение библиотеки numpy для операция над тензорами и решения задач линейной алгебры и генерации тензоров различной размерности по заданным правилам

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.

2. 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

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

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

Предварительная обработка данных при решении задач анализа данных

Методы предварительной обработки: заполнение пропусков в данных, обработка выбросов в данных.

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.

3. 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

Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных

Приемы отбора лучших моделей для решения поставленных задач. Различные методики разбиения выборок для более стабильной и правдоподобной оценки

Методы подбора гипер параметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных

Инструменты построения конвейеров обработки данных в машинном обучении. Методы и инструменты автоматического подбора гипер параметров моделей машинного обучения, а также подбора самих моделей и методов предварительной обработки

4. 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.

5. Задачи обучения без учителя: понижение размерности

Обзор задач обучения без учителя и их применимость на практике. Методы понижения размерности и их практические приложения

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.

Задачи обучения без учителя: поиск аномалий

Постановка задачи поиска аномалий и выявления новизны. Методы поиска аномалий на основе подходов машинного обучения.

6. Обработка естественного языка: предварительная обработка

Знакомство с задачами обработки естественного языка.: Методы первичной предварительной обработки текстов: очистка, токенизация, лемматизация, стемминг.

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.

Обработка естественного языка: типы решаемых задач

Рассмотрение спектра задач обработки естественного языка: моделирование языка, классификация, поиск именованных сущностей, суммаризация, генерация.

7. Введение в нейронные сети

Базовые понятия нейронных сетей: нейрон, обучение нейрона, нейронная сеть, обучение нейронной сети. Функции потерь и метрики качества.

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.

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.

Нейронные сети для решения задач обработки изображений

Знакомство с задачами обработки изображений. Изучение сверточных нейронных сетей. Понятие ядра, свертки и подвыборки.

Нейронные сети для решения задач обработки изображений: современные архитектуры

Изучение современных архитектур нейронных сетей для решения задач обработки изображений: VGG-16, VGG-19, ResNet, Inception, Xception, DenseNet, MobileNet, EfficientNet.

8. Нейронные сети для решения задач обработки естественного языка

Архитектуры нейронных сетей для решения задач обработки естественного языка: одномерные сверточные сети, рекуррентные нейронные сети.

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

9. 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

Содержание лабораторных занятий по дисциплине

1. Обзор инструментов, используемых 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 гипер параметров 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
10. Обработка естественного языка: предварительная обработка
11. Обработка естественного языка: векторные модели
12. Введение в нейронные сети. Нейронные сети для решения задач регрессии
13. Нейронные сети для решения задач классификации
14. Нейронные сети для решения задач обработки изображений
15. Нейронные сети для решения задач обработки изображений: современные архитектуры
16. Нейронные сети для решения задач обработки естественного языка
17. Нейронные сети для решения задач обработки естественного языка: современные архитектуры
18. Нейронные сети для решения задач предсказания временных рядов

5. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

5.1. Текущий контроль успеваемости

Рейтинг-контроль №1

1. Обзор инструментов, применяемых для анализа данных и машинного обучения.
2. Обзор библиотеки NumPy
3. Работа с библиотекой Pandas (Matplotlib, seaborn)

4. Модели машинного обучения для решения задач классификации
5. Модели машинного обучения для решения задач регрессии
6. Метрики качества при решении задач классификации и регрессии
7. Предварительная обработка данных при решении задач анализа данных
8. Преобразование и создание новых признаков для решения задач анализа данных
9. Методы выбора лучших признаков для решения задач классификации и регрессии
10. Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных
11. Методы подбора гипер параметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных

Рейтинг-контроль №2

1. Работа с временными рядами
2. Задачи обучения без учителя: понижение размерности
3. Задачи обучения без учителя: кластеризация
4. Задачи обучения без учителя: поиск аномалий
5. Обработка естественного языка: предварительная обработка
6. Обработка естественного языка: векторные модели
7. Обработка естественного языка: типы решаемых задач

Рейтинг-контроль №3

1. Введение в нейронные сети
2. Нейронные сети для решения задач регрессии
3. Нейронные сети для решения задач классификации
4. Нейронные сети для решения задач обработки изображений
5. Нейронные сети для решения задач обработки изображений: современные архитектуры
6. Нейронные сети для решения задач обработки естественного языка
7. Нейронные сети для решения задач обработки естественного языка: современные архитектуры
8. Нейронные сети для решения задач предсказания временных рядов

5.2. Промежуточная аттестация по итогам освоения дисциплины (зачет с оценкой)

Список примерных вопросов для Зачета с оценкой:

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.
31. Опишите устройство искусственного нейрона
32. Опишите процесс обучения искусственного нейрона
33. Опишите функции активации, которые вы знаете. Проведите их сравнительный анализ
34. Опишите устройство полносвязной нейронной сети.
35. Какие задачи может решать нейронная сеть? Как изменяется выходной слой и функция активации на выходном слое в зависимости от решаемой задачи?
36. Опишите процесс обучения нейронной сети. На чем основан метод обратного распространения ошибки?
37. Расскажите про метрики качества нейронных сетей для решения различных задач. Проведите сравнительный анализ
38. Расскажите про функции потерь нейронных сетей для решения различных задач. Проведите сравнительный анализ
39. Что такое переобучение нейронной сети? Какие методы борьбы с переобучением вы знаете? Как можно выявить факт переобучения?
40. Какие гипер параметры нейронной сети вы знаете? На что они влияют? Как происходит процесс подбора гипер параметров?

41. Что такое градиентный спуск? Опишите как работает градиентный спуск и стохастический градиентный спуск.
42. Опишите необходимые свойства функций ошибок, для которых можно применить градиентный спуск.
43. Какие модификации градиентного спуска вы знаете? Проведите сравнительный анализ модификаций градиентного спуска
44. Опишите недостатками полносвязных нейронных сетей для решения задач обработки изображений.
45. Опишите устройство сверточной нейронной сети.
46. Опишите принцип работы сверточного слоя.
47. Для чего нужны ядра в сверточном слое? Как они подбираются?
48. Как происходит процесс переноса обучения в сверточных нейронных сетях?
49. Какие современные архитектуры нейронных сетей для решения задач обработки изображений вы знаете? Проведите краткий сравнительный анализ.
50. Опишите принцип работы нейронной сети с архитектурой ResNet
51. Опишите принцип работы нейронной сети с архитектурой Inception
52. Опишите принцип работы нейронной сети с архитектурой Xception
53. Опишите принцип работы нейронной сети с архитектурой VGG-16/19
54. Опишите принцип работы рекуррентных нейронных сетей? Какие задачи они решают?
55. Опишите принцип работы LSTM блока в нейронной сети.
56. Опишите принцип работы GRU блока в нейронной сети.
57. Опишите принцип работы двунаправленных рекуррентных нейронных сетей. Для чего они применяются?
58. Какие задачи решают нейронные сети при работе с изображениями?
59. Какие задачи решают нейронные сети при работе с текстовой информацией?
60. Какие векторные модели текстов вы знаете? Проведите их сравнительный анализ
61. Что такое word2vec? Опишите процесс его обучения и особенности применения.

5.3. Самостоятельная работа обучающегося

Самостоятельная работа обучающихся заключается в самостоятельном изучении отдельных тем, практической реализации заданий контрольных и самостоятельных работ по этим темам. Контроль выполнения самостоятельной работы проводится при текущих контрольных мероприятиях и на промежуточной аттестации по итогам освоения дисциплины. Учебно-методическое обеспечение самостоятельной работы – основная литература [1-3].

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 на платформу курса.

Примерная тематика СРС:

Самостоятельная 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

Самостоятельная 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 50к, имеют как минимум высшее образование? (признак education – Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters или Doctorate)
5. Среди кого больше доля зарабатывающих много (>50К): 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?

Самостоятельная работа №3:

Дан colab 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.

Самостоятельная работа №4:

Дан colab 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.

Фонд оценочных материалов (ФОМ) для проведения аттестации уровня сформированности компетенций обучающихся по дисциплине оформляется отдельным документом.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Книгообеспеченность

Наименование литературы: автор, название, вид издания, издательство	Год издания	КНИГООБЕСПЕЧЕННОСТЬ
		Наличие в электронном каталоге ЭБС
Основная литература		
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.	2021	https://e.lanbook.com/book/173811
2. Тарасов, И. Е. Статистический анализ	2020	https://e.lanbook.com/book/163854

данных в информационных системах : учебно-методическое пособие / И. Е. Тарасов. — Москва : РТУ МИРЭА, 2020. — 96 с. — Текст : электронный // Лань : электронно-библиотечная система		
3. 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	2019	https://e.lanbook.com/book/122180
Дополнительная литература		
1. 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	2019	https://e.lanbook.com/book/131691

6.2. Периодические издания

1. Вестник компьютерных и информационных технологий ISSN 1810-7206.
2. Цифровая библиотека научно-технических изданий Института инженеров по электротехнике и радиоэлектронике (Institute of Electrical and Electronic Engineers (IEEE)) на английском языке – <http://www.ieee.org/ieeexplore>

6.3. Интернет-ресурсы

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. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
5. Университетская библиотека ONLINE – <https://biblioclub.ru/>
6. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
7. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
8. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>
9. Портал российского образования www.edu.ru
10. Портал российских электронных библиотек www.elbib.ru
11. Научная электронная библиотека www.eLibrary.ru
12. Научная библиотека ВлГУ library.vlsu.ru
13. Учебный сайт кафедры ИСПИ ВлГУ <https://ispi.cdo.vlsu.ru>
14. Электронная библиотечная система ВлГУ <https://vlsu.bibliotech.ru/>
15. Web of Science Core Collection – <http://apps.webofknowledge.com/>

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для реализации данной дисциплины имеются специальные помещения для проведения занятий: занятий лекционного типа, занятий лабораторного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы. Лабораторные работы проводятся в компьютерном классе, оборудованном мультимедийным проектором с экраном и обеспеченным доступом в Интернет.

Перечень используемого лицензионного программного обеспечения:

- Операционная система Microsoft Windows 10
- Офисный пакет Microsoft Office 2016 (MS PowerPoint, MS Word, MS Excel, MS Teams)
- Бесплатно-распространяемое программное обеспечение Python –

<https://www.python.org/>

Рабочую программу составил: зав. каф. ИСПИ И.Е. Жигалов 

Рецензент: к.т.н., ведущий специалист отдела ИТ ООО «Дау Изолан» Фадин Д.Н. 

Программа рассмотрена и одобрена на заседании кафедры ИСПИ

Протокол № 5 от 15.12.21 года

Заведующий кафедрой И.Е. Жигалов 

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии направления 09.04.04 «Программная инженерия»

Протокол № 5 от 15.12.21 года

Председатель комиссии И.Е. Жигалов 

**ЛИСТ ПЕРЕУТВЕРЖДЕНИЯ
РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ**

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Рабочая программа одобрена на 20____ / 20____ учебный года

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой _____

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)**

Институт информационных технологий и радиоэлектроники
Кафедра информационных систем и программной инженерии

УТВЕРЖДАЮ
Заведующий кафедрой


И.Е. Жигалов

« 15 » 12 20 21

Основание:
решение кафедры

от « 15 » 12 20 21

**ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ
ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ПРИ ИЗУЧЕНИИ УЧЕБНОЙ ДИСЦИПЛИНЫ
«Спортивный анализ данных»**

Направление подготовки: 09.04.04 «Программная инженерия»

Профиль подготовки: Инженерия искусственного интеллекта

Уровень высшего образования: магистратура

Владимир, 2021 г.

ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ МАТЕРИАЛОВ

Фонд оценочных материалов (ФОМ) для текущего контроля успеваемости и промежуточной аттестации по дисциплине «Спортивный анализ данных» разработан в соответствии с рабочей программой, входящей в ОПОП направления подготовки 09.04.04 «Программная инженерия», профиль подготовки «Инженерия искусственного интеллекта».

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного материала
1	Обзор инструментов, применяемых для анализа данных и машинного обучения. Обзор библиотеки Numpy Работа с библиотекой Pandas (Matplotlib, seaborn)	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
2	Модели машинного обучения для решения задач классификации Модели машинного обучения для решения задач регрессии Метрики качества при решении задач классификации и регрессии Предварительная обработка данных при решении задач анализа данных Преобразование и создание новых признаков для решения задач анализа данных	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
3	Методы выбора лучших признаков для решения задач классификации и регрессии Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных Методы подбора гиперпараметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
4	Работа с временными рядами	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой

5	Задачи обучения без учителя: понижение размерности Задачи обучения без учителя: кластеризация Задачи обучения без учителя: поиск аномалий	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
6	Обработка естественного языка: предварительная обработка Обработка естественного языка: векторные модели Обработка естественного языка: типы решаемых задач	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
7	Введение в нейронные сети Нейронные сети для решения задач регрессии Нейронные сети для решения задач классификации Нейронные сети для решения задач обработки изображений Нейронные сети для решения задач обработки изображений: современные архитектуры	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
8	Нейронные сети для решения задач обработки естественного языка Нейронные сети для решения задач обработки естественного языка: современные архитектуры	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой
9	Нейронные сети для решения задач предсказания временных рядов	ПК-7	Контрольные работы, выполнение лабораторных работ, рейтинг-контроль, зачет с оценкой

Комплект оценочных материалов по дисциплине «Спортивный анализ данных» предназначен для аттестации обучающихся на соответствие их персональных достижений поэтапным требованиям образовательной программы, в том числе рабочей программы дисциплины «Спортивный анализ данных», для оценивания результатов обучения: знаний, умений, владений и уровня приобретенных компетенций.

Комплект оценочных материалов по дисциплине «Спортивный анализ данных» включает:

1. Оценочные материалы для проведения текущего контроля успеваемости:
 - комплект вопросов рейтинг-контроля, позволяющих оценивать знание фактического материала (базовые понятия, алгоритмы, факты) и умение правильно использовать специальные термины и понятия, распознавание объектов изучения в рамках определенного раздела дисциплины;

- комплект вопросов для контроля самостоятельной работы обучающихся, позволяющих оценивать знание фактического материала.

2. Оценочные материалы для проведения промежуточной аттестации в форме

- контрольные вопросы для проведения зачета с оценкой, позволяющие провести процедуру измерения уровня знаний и умений обучающихся.

Перечень компетенций, формируемых в процессе изучения дисциплины «Спортивный анализ данных» при освоении образовательной программы по направлению подготовки 09.04.04 «Программная инженерия»

<i>ПК-7. Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях</i>		
<i>Знать</i>	<i>Уметь</i>	<i>Иметь навыки</i>
принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»; современное состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта	руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	проведения анализа новых направлений, методов и технологий в области искусственного интеллекта и определения наиболее перспективных для различных областей применения

Показатели, критерии и шкала оценивания компетенций текущего контроля знаний по учебной дисциплине «Спортивный анализ данных»

Текущий контроль знаний, согласно «Положению о рейтинговой системе комплексной оценки знаний студентов в ВлГУ» (далее Положение) в рамках изучения дисциплины «Спортивный анализ данных» предполагает письменный рейтинг-контроль, выполнение и защиту лабораторных работ, компьютерное тестирование.

Регламент проведения письменного рейтинг-контроля

№	Вид работы	Продолжительность
1	Предел длительности рейтинг-контроля	35-40 мин.
2	Внесение исправлений	до 5 мин.
	Итого	до 45 мин.

Критерии оценки письменного рейтинг-контроля

Результаты каждого письменного рейтинга оцениваются в баллах. Максимальная сумма, набираемая студентом на каждом письменном рейтинге, составляет 10 баллов.

Критерии оценки для письменного рейтинга:

- 9-10 баллов выставляется обучающемуся, если соблюдаются критерии: полное раскрытие темы, вопроса, указание точных названий и определений, правильная формулировка понятий и категорий, приведение формул и (в необходимых случаях) их

вывода, приведение статистики, самостоятельность ответа, использование дополнительной литературы;

- 7-8 баллов выставляется обучающемуся, если соблюдаются критерии: недостаточно полное раскрытие темы, несущественные ошибки в определении понятий и категорий, формулах, выводе формул, статистических данных, кардинально не меняющих суть изложения, наличие грамматических и стилистических ошибок, использование устаревшей учебной литературы;

- 6-7 баллов выставляется обучающемуся, если соблюдаются критерии: отражение лишь общего направления изложения лекционного материала и материала современных учебников, наличие достаточно количества несущественных или одной-двух существенных ошибок в определении понятий и категорий, формулах, их выводе, статистических данных, наличие грамматических и стилистических ошибок, использование устаревшей учебной литературы, неспособность осветить проблематику дисциплины;

- 1-6 выставляется обучающемуся, если соблюдаются критерии: нераскрытые темы; большое количество существенных ошибок, наличие грамматических и стилистических ошибок, отсутствие необходимых умений и навыков.

Регламент проведения лабораторных работ

В целях закрепления практического материала и углубления теоретических знаний по разделам дисциплины «Спортивный анализ данных» предполагается выполнение лабораторных работ, что позволяет углубить процесс познания, раскрыть понимание прикладной значимости осваиваемой дисциплины.

Лабораторные работы выполняются на компьютерах.

Критерии оценки выполнения лабораторных работ

Результаты выполнения каждой лабораторной работы оцениваются в баллах. Максимальная сумма, набираемая студентом за выполнение каждой лабораторной работы, составляет 1 балл.

Критерии оценки для выполнения лабораторной работы:

- 0,9-1 балл выставляется обучающемуся, если соблюдаются критерии: представлен полный письменный отчет по лабораторной работе, содержащий описание всех этапов ее выполнения и надлежащим образом оформленный (в печатном или электронном виде - в соответствии с требованием преподавателя), полностью выполнено задание на лабораторную работу, обучающийся верно и полно ответил на все контрольные вопросы преподавателя по теоретической и практической части лабораторной работы, лабораторная работа выполнена самостоятельно и в определенный преподавателем срок;

- 0,7-0,8 баллов выставляется обучающемуся, если соблюдаются критерии: представлен недостаточно полный письменный отчет по лабораторной работе, содержащий описание всех этапов ее выполнения, имеющий, возможно, погрешности в оформлении (в печатном или электронном виде - в соответствии с требованием преподавателя), полностью выполнено задание на лабораторную работу, обучающийся преимущественно верно и полно ответил на контрольные вопросы преподавателя по теоретической и практической части лабораторной работы, лабораторная работа выполнена самостоятельно, возможно, с нарушением определенного преподавателем срока предоставления отчета, отчет содержит грамматические и стилистические ошибки;

- 0,6-0,7 баллов выставляется обучающемуся, если соблюдаются критерии: представлен недостаточно полный письменный отчет по лабораторной работе, содержащий описание не всех этапов ее выполнения, имеющий, возможно, погрешности в оформлении (в печатном или электронном виде - в соответствии с требованием преподавателя), в основном выполнено задание на лабораторную работу, обучающийся ответил на контрольные вопросы преподавателя по теоретической и практической части лабораторной работы с отражением

лишь общего направления изложения материала, с наличием достаточно количества несущественных или одной-двух существенных ошибок, лабораторная работа выполнена самостоятельно, с нарушением определенного преподавателем срока предоставления отчета, отчет содержит грамматические и стилистические ошибки, при его составлении использована устаревшая учебная литература;

- 0,1-0,6 выставляется обучающемуся, если соблюдаются критерии: письменный отчет по лабораторной работе (в печатном или электронном виде - в соответствии с требованием преподавателя) не представлен или представлен неполный, отчет содержит описание не всех этапов выполнения работы, имеет погрешности в оформлении, задание на лабораторную работу выполнено не полностью, обучающийся ответил на контрольные вопросы преподавателя по теоретической и практической части лабораторной работы с большим количеством существенных ошибок, продемонстрировал неспособность осветить проблематику лабораторной работы, лабораторная работа выполнена несамостоятельно, с существенным нарушением определенного преподавателем срока предоставления отчета, отчет содержит грамматические и стилистические ошибки, при его составлении использована устаревшая учебная литература, обучающийся при выполнении работы продемонстрировал отсутствие необходимых умений и практических навыков.

При оценке за лабораторную работу менее 0,6 баллов, данная работа считается невыполненной и не зачитывается. При невыполнении лабораторной работы хотя бы по одной из изучаемых тем, обучающийся не получает положительную оценку при промежуточном контроле по дисциплине (зачете с оценкой).

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ ЗНАНИЙ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ «Спортивный анализ данных»

Перечень вопросов для текущего контроля знаний (письменный рейтинг-контроль)

Рейтинг-контроль №1

1. Обзор инструментов, применяемых для анализа данных и машинного обучения.
2. Обзор библиотеки Numpy
3. Работа с библиотекой Pandas (Matplotlib, seaborn)
4. Модели машинного обучения для решения задач классификации
5. Модели машинного обучения для решения задач регрессии
6. Метрики качества при решении задач классификации и регрессии
7. Предварительная обработка данных при решении задач анализа данных
8. Преобразование и создание новых признаков для решения задач анализа данных
9. Методы выбора лучших признаков для решения задач классификации и регрессии
10. Методы выбора лучших моделей и алгоритмов для решения поставленных задач анализа данных
11. Методы подбора гипер параметров моделей машинного обучения и построения конвейеров автоматизации решения задач анализа данных

Рейтинг-контроль №2

1. Работа с временными рядами
2. Задачи обучения без учителя: понижение размерности
3. Задачи обучения без учителя: кластеризация
4. Задачи обучения без учителя: поиск аномалий
5. Обработка естественного языка: предварительная обработка
6. Обработка естественного языка: векторные модели

7. Обработка естественного языка: типы решаемых задач

Рейтинг-контроль №3

1. Введение в нейронные сети
2. Нейронные сети для решения задач регрессии
3. Нейронные сети для решения задач классификации
4. Нейронные сети для решения задач обработки изображений
5. Нейронные сети для решения задач обработки изображений: современные архитектуры
6. Нейронные сети для решения задач обработки естественного языка
7. Нейронные сети для решения задач обработки естественного языка: современные архитектуры
8. Нейронные сети для решения задач предсказания временных рядов

Темы лабораторных работ:

1. Обзор инструментов, используемых 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 гипер параметров 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
10. Обработка естественного языка: предварительная обработка
11. Обработка естественного языка: векторные модели
12. Введение в нейронные сети. Нейронные сети для решения задач регрессии
13. Нейронные сети для решения задач классификации
14. Нейронные сети для решения задач обработки изображений
15. Нейронные сети для решения задач обработки изображений: современные архитектуры
16. Нейронные сети для решения задач обработки естественного языка
17. Нейронные сети для решения задач обработки естественного языка: современные архитектуры
18. Нейронные сети для решения задач предсказания временных рядов

Перечень вопросов для контроля самостоятельной работы обучающегося

Самостоятельная работа обучающихся заключается в самостоятельном изучении отдельных тем, практической реализации заданий контрольных и самостоятельных работ по этим темам. Контроль выполнения самостоятельной работы проводится при текущих контрольных мероприятиях и на промежуточной аттестации по итогам освоения дисциплины.

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 на платформу курса.

Примерная тематика СРС:

Самостоятельная работа №1:

Дан colaboratory 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

Самостоятельная 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?

Самостоятельная работа №3:

Дан colab 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.

Самостоятельная работа №4:

Дан colab 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.

Общее распределение баллов текущего и промежуточного контроля по видам учебных работ для студентов (в соответствии с Положением)

№	Пункт	Максимальное число баллов
1	Письменный рейтинг-контроль 1	10
2	Письменный рейтинг-контроль 2	10
3	Письменный рейтинг-контроль 3	10
4	Посещение занятий студентом	5
5	Дополнительные баллы (бонусы)	5
6	Выполнение лабораторных работ и семестрового плана самостоятельной работы	60
8	Всего	100

Показатели, критерии и шкала оценивания компетенций промежуточной аттестации знаний по учебной дисциплине «Спортивный анализ данных» на зачете с оценкой

Регламент проведения промежуточного контроля (зачета с оценкой)

Промежуточная аттестация по итогам освоения дисциплины (зачет с оценкой) проводится перед экзаменационной сессией. Зачет проставляется студенту после выполнения студентом семестрового плана самостоятельной работы.

Критерии оценивания компетенций при проставлении зачета

Критерии оценки для промежуточного контроля (зачета с оценкой):

- оценка «отлично» (соответствует 91-100 баллов по шкале рейтинга) выставляется обучающемуся, если соблюдаются критерии: теоретическое содержание оцениваемой части дисциплины освоено полностью, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены в установленные сроки, качество их выполнения оценено числом баллов, близким к максимальному;

- оценка «хорошо» (соответствует 74-90 баллов по шкале рейтинга) выставляется обучающемуся, если соблюдаются критерии: теоретическое содержание курса освоено полностью, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками или с нарушением установленных сроков;

- оценка «удовлетворительно» (соответствует 61-73 баллов по шкале рейтинга) выставляется обучающемуся, если соблюдаются критерии: теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки;

- оценка «неудовлетворительно» (соответствует менее 60 баллов по шкале рейтинга) выставляется обучающемуся, если соблюдаются критерии: теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, выполненные учебные задания содержат грубые ошибки.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ «Спортивный анализ данных»

Перечень вопросов для промежуточной аттестации (зачет с оценкой)

Список примерных вопросов для Зачета с оценкой:

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.
31. Опишите устройство искусственного нейрона
32. Опишите процесс обучения искусственного нейрона
33. Опишите функции активации, которые вы знаете. Проведите их сравнительный анализ
34. Опишите устройство полносвязной нейронной сети.
35. Какие задачи может решать нейронная сеть? Как изменяется выходной слой и функция активации на выходном слое в зависимости от решаемой задачи?
36. Опишите процесс обучения нейронной сети. На чем основан метод обратного распространения ошибки?
37. Расскажите про метрики качества нейронных сетей для решения различных задач. Проведите сравнительный анализ
38. Расскажите про функции потерь нейронных сетей для решения различных задач. Проведите сравнительный анализ
39. Что такое переобучение нейронной сети? Какие методы борьбы с переобучением вы знаете? Как можно выявить факт переобучения?
40. Какие гипер параметры нейронной сети вы знаете? На что они влияют? Как происходит процесс подбора гипер параметров?
41. Что такое градиентный спуск? Опишите как работает градиентный спуск и стохастический градиентный спуск.
42. Опишите необходимые свойства функций ошибок, для которых можно применить градиентный спуск.
43. Какие модификации градиентного спуска вы знаете? Проведите сравнительный анализ модификаций градиентного спуска
44. Опишите недостатками полносвязных нейронных сетей для решения задач обработки изображений.
45. Опишите устройство сверточной нейронной сети.
46. Опишите принцип работы сверточного слоя.
47. Для чего нужны ядра в сверточном слое? Как они подбираются?
48. Как происходит процесс переноса обучения в сверточных нейронных сетях?
49. Какие современные архитектуры нейронных сетей для решения задач обработки изображений вы знаете? Проведите краткий сравнительный анализ.
50. Опишите принцип работы нейронной сети с архитектурой ResNet
51. Опишите принцип работы нейронной сети с архитектурой Inception
52. Опишите принцип работы нейронной сети с архитектурой Xception
53. Опишите принцип работы нейронной сети с архитектурой VGG-16/19
54. Опишите принцип работы рекуррентных нейронных сетей? Какие задачи они решают?
55. Опишите принцип работы LSTM блока в нейронной сети.
56. Опишите принцип работы GRU блока в нейронной сети.
57. Опишите принцип работы двунаправленных рекуррентных нейронных сетей. Для чего они применяются?
58. Какие задачи решают нейронные сети при работе с изображениями?
59. Какие задачи решают нейронные сети при работе с текстовой информацией?
60. Какие векторные модели текстов вы знаете? Проведите их сравнительный анализ
61. Что такое word2vec? Опишите процесс его обучения и особенности применения.

Критерии оценивания компетенций при аттестации по дисциплине

Максимальная сумма баллов, набираемая студентом по дисциплине «Спортивный анализ данных» в течение семестра равна 100

Оценка в баллах	Оценка по дисциплине	Критерии оценивания компетенций	Уровень сформированности компетенций
91 - 100	«Отлично»	Теоретическое содержание курса освоено полностью без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.	Высокий
74 - 90	«Хорошо»	Теоретическое содержание курса освоено полностью без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.	Продвинутый
61 - 73	«Удовлетворительно»	Теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.	Пороговый
0 - 60	«Неудовлетворительно»	Теоретическое содержание курса не освоено, необходимые практические навыки работы не сформированы, выполненные учебные задания содержат грубые ошибки.	Компетенции не сформированы