



КГЭУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО «КГЭУ»)

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ**  
**для проведения текущего контроля успеваемости и промежуточной**  
**аттестации студентов по итогам освоения дисциплины**

**Спортивный анализ данных**

Направление подготовки      09.04.01 Информатика и вычислительная техника

Направленность (профиль)    Инженерия искусственного интеллекта

Квалификация                    Магистр

Форма обучения                 Очная

Разработчик

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

Оценочные материалы оформлены в соответствии с ПОЛОЖЕНИЕМ О ПОРЯДКЕ РАЗРАБОТКИ И УТВЕРЖДЕНИЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ – ПРОГРАММ БАКАЛАВРИАТА, ПРОГРАММ СПЕЦИАЛИТЕТА И ПРОГРАММ МАГИСТРАТУРЫ В КГЭУ

г. Казань, 2021

## 1. Цель и задачи текущего контроля и промежуточной аттестации студентов по дисциплине «Спортивный анализ данных»

*Цель текущего контроля* - систематическая проверка степени освоения программы дисциплины «Спортивный анализ данных», уровня сформированности знаний, умений, навыков, компетенций на текущих занятиях

*Задачи текущего контроля:*

1. определение индивидуального учебного рейтинга студентов;
2. своевременное выполнение корректирующих действий по содержанию и организации процесса обучения; обнаружение и устранение пробелов в усвоении учебной дисциплины;
3. подготовки к промежуточной аттестации.

В течение семестра при изучении дисциплины реализуется комплексная система поэтапного оценивания уровня освоения – бально-рейтинговая система. За каждый вид учебных действий студенты получают определенное количество баллов. В течение семестра студент может набрать до 60-ти баллов.

*Цель промежуточной аттестации* - проверка степени усвоения студентами учебного материала за время изучения дисциплины, уровня сформированности компетенций после завершения изучения дисциплины. Аттестация проходит в форме зачета.

*Задачи промежуточной аттестации:*

1. определение уровня усвоения учебной дисциплины;
2. определение уровня сформированности компетенций.

## 2. Основное содержание текущего контроля и промежуточной аттестации студентов

В результате изучения дисциплины «Спортивный анализ данных» формируются следующие компетенции или их составляющие:

### 2.1. Основное содержание текущего контроля

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Содержание оценочных заданий для выявления сформированности компетенций		
		Базовый уровень	Продвинутый уровень	Высокий уровень
ПК-6	ПК-6.1. Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях	Решение практического задания	Решение практического задания, написание отчета	Решение практического задания, написание отчета, ответы на вопросы

### 2.2. Основное содержание промежуточной аттестации студентов

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Содержание оценочных заданий для выявления сформированности компетенций		
		Базовый уровень	Продвинутый уровень	Высокий уровень
ПК-6	ПК-6.1. Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях	Решение кейса	Решение кейса	Решение кейса

### 3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

п/п	Наименование компоненты	Критерии оценки	Максимальное число баллов
1	Практические занятия - 4 занятия	выполнение задания 8 баллов оформление выводов 4 балла ответы на вопросы 3 балла	15 баллов за каждое занятие <b>60 баллов</b>
2	Решение кейса	раскрыты теоретические предпосылки модели 10 баллов проведено моделирование, получены оценки параметров модели 20 баллов проведена оценка пригодности модели и оценка значимости коэффициентов 10 баллов	<b>40 баллов</b>
		ИТОГО	<b>100 баллов</b>

#### Содержание практических заданий

##### Практическое задание 1. Задачи обучения без учителя: понижение размерности, кластеризация

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Используя метод главных компонент, убедитесь в целесообразности снижения размерности, выберите оптимальное количество главных компонент
4. Рассчитать главные компоненты и дать им характеристику.
5. Используя полученные главные компоненты, провести иерархическую кластеризацию и построить дендрограмму. Определить оптимальное количество кластеров.
6. Используя имеющиеся в Scikit-Learn методы кластеризации, классифицировать объекты различными методами.
7. Проанализировать результаты

##### Практическое задание 2. Модели машинного обучения для решения задач регрессии

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Провести корреляционный анализ. Выявить коллинеарные факторы.
4. Используя библиотеку *statsmodels* и метод пошагового исключения, отобрать факторы в модель (не забудьте про свободный член!!). Учесть результаты корреляционного анализа.
5. Разделить выборку на тренировочную и тестовую случайным(!) образом.
6. Обучить модель на тренировочной выборке.
7. Рассчитать значение ошибок MSE и RMSE и  $R^2$  для тренировочных данных:
8. Рассчитать значение ошибок MSE и RMSE и  $R^2$  для тестовых данных:

Повторить шаги 5-7 несколько раз (отдельный код писать не надо – просто несколько раз запустить). Что можно сказать по поводу построенной модели?

9. Какие факторы вошли в модель (словами, а не именами переменных)?
10. Записать модель
11. Построить аналогичную модель методом градиентного спуска.
12. Оценить, какой метод более эффективен.
13. Дать экономическую интерпретацию коэффициентам регрессии.
14. Проанализировать результаты

### **Практическое задание 3. Модели машинного обучения для решения задач классификации**

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Построить модель логистической регрессии. Учтите, что в модели должны остаться только значимые факторы.
4. Какие факторы вошли в модель (словами, а не именами переменных)?
5. Записать модель
6. Получить расчетные значения
7. Проанализировать метрики качества модели
8. Дать экономическую интерпретацию коэффициентам регрессии.
9. Проанализировать результаты

### **Практическое задание 4. Нейронные сети**

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Скопируйте базовое решение и запустите его у себя.
2. Изменяя любые параметры архитектуры нейронной сети или параметры обучения нейронной сети достигнуть наибольшего качества модели.
3. Ваша цель достигнуть качества модели на тестовых данных 0.9 и выше.

### ***Вопросы для подготовки к зачету по курсу «Анализ временных рядов»***

1. Назовите ключевые типы данных, которые могут быть в задачах интеллектуального анализа данных.
3. Назовите задачи, которые могут решаться при обучении моделей машинного обучения с учителем. Приведите примеры таких задач.
4. Назовите основные этапы решения задачи интеллектуального анализа данных.
5. Опишите, какую предварительную обработку и подготовку данных необходимо провести при работе с задачами на изображениях.
6. Какие инструменты и модели подходят для работы с изображениями?
7. Опишите, какую предварительную обработку и подготовку данных необходимо провести при работе с задачами на табличных данных.
8. Какие инструменты и модели подходят для работы с табличными данными?
9. Опишите ключевые метрики качества моделей машинного обучения при решении задачи классификации. Какие из них не чувствительны к несбалансированным данным?
10. Опишите ключевые метрики качества моделей машинного обучения при решении задачи регрессии
11. Опишите процесс выбора признаков, которые положительно влияют на качество итоговой модели машинного обучения.

12. Для чего нужна оценка качества модели. Какие подходы к проверке качества моделей вы знаете? Назовите преимущества и недостатки основных подходов.
13. Опишите устройство искусственного нейрона
14. Опишите процесс обучения искусственного нейрона
15. Опишите функции активации, которые вы знаете. Проведите их сравнительный анализ
16. Опишите устройство полносвязной нейронной сети.
17. Какие задачи может решать нейронная сеть? Как изменяется выходной слой и функция активации на выходном слое в зависимости от решаемой задачи?
18. Опишите процесс обучения нейронной сети. На чем основан метод обратного распространения ошибки?
19. Расскажите про метрики качества нейронных сетей для решения различных задач. Проведите сравнительный анализ
20. Расскажите про функции потерь нейронных сетей для решения различных задач. Проведите сравнительный анализ
21. Что такое переобучение нейронной сети? Какие методы борьбы с переобучением вы знаете? Как можно выявить факт переобучения?
22. Какие гипер параметры нейронной сети вы знаете? На что они влияют? Как происходит процесс подбора гипер параметров?
23. Что такое градиентный спуск? Опишите как работает градиентный спуск и стохастический градиентный спуск.
24. Опишите необходимые свойства функций ошибок, для которых можно применить градиентный спуск.
25. Какие модификации градиентного спуска вы знаете? Проведите сравнительный анализ модификаций градиентного спуска
26. Опишите недостатки полносвязных нейронных сетей для решения задач обработки изображений.
27. Опишите устройство сверточной нейронной сети.
28. Опишите принцип работы сверточного слоя.
29. Для чего нужны ядра в сверточном слое? Как они подбираются?
30. Как происходит процесс переноса обучения в сверточных нейронных сетях?
31. Какие современные архитектуры нейронных сетей для решения задач обработки изображений вы знаете? Проведите краткий сравнительный анализ.

### ***Пример типового кейса для проведения промежуточной аттестации***

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Сделать предположение о подходящем методе (методах) моделирования данных
2. Построить модель, получить оценки параметров модели
3. Оценить качество/пригодность модели
4. Получить расчетные значения
5. Построить графики исходных и расчетных данных (на одном графике)