

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Писарев Сергей Станиславович
Должность: Ректор
Дата подписания: 24.09.2024 12:35:18
Уникальный программный ключ:
b9d7463b91f434da3d4dc1afa9a0cf32d3c58650

**Негосударственное образовательное учреждение высшего образования
«Школа управления СКОЛКОВО»**

Утверждено
Ректор С.С. Писарев
29 августа 2024 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Наука о данных**

Направление подготовки	38.03.02 Менеджмент
Квалификация выпускника	Бакалавр
Образовательная программа	Управление и предпринимательство
Форма обучения	Очная
Рабочая программа дисциплины разработана	

Трудоемкость		Контактная работа		Самостоятельная работа	Форма контроля	Семестр/кв артиль
з.е.	часы	лекции	семинарские занятия			
5	180	24	24	132	Экзамен	3

**Москва
2024**

1. АННОТАЦИЯ ДИСЦИПЛИНЫ

В рамках дисциплины “Наука о данных” студенты получают знания и навыки, необходимые для постановки задачи анализа данных; предварительной обработки данных; разработки, реализации и применения методов интеллектуального анализа данных в профессиональной деятельности.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В случае успешного освоения курса студенты будут:

знать

- основные положения науки о данных;
- основные категории данных;
- методы работы с данными;
- способы визуализации данных

уметь

- формулировать бизнес-задачи в терминологии анализа данных;
- применять методы интеллектуального анализа данных к анализу и использованию больших данных;
- принимать управленческие решения на основе данных

владеть

- навыком сбора и анализа данных;
- навыком визуализации данных;
- информационно-аналитическими системами для работы с данными;
- навыком интеграции данных из разных источников, их интерпретации в контексте поставленной задачи;
- навыками применения технологии MapReduce и ее реализации Hadoop

Дисциплина направлена на развитие следующих компетенций и их индикаторов:

Код компетенции	Формулировка компетенции и/или ее индикатора (ов)
ОПК-2.	Способен осуществлять сбор, обработку и анализ данных, необходимых для решения поставленных управленческих задач, с использованием современного инструментария и интеллектуальных информационно-аналитических систем
ОПК-2-1.	Определяет источники данных и выбирает методы и инструменты поиска, корректно осуществляет анализ литературы и документов
ОПК-2-2.	Применяет методы сбора, обработки и анализа данных, необходимых для решения управленческих задач, с использованием современных цифровых технологий, воспринимает, анализирует, запоминает и передает информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными.
ОПК-5.	Способен использовать при решении профессиональных задач современные информационные технологии и программные

	средства, включая управление крупными массивами данных и их интеллектуальный анализ
ОПК-5-1.	Использует цифровые технологии, включая информационные системы и базы данных, системы искусственного интеллекта и системы интеллектуального анализа и обработки данных для решения профессиональных задач
ОПК-5-2.	Оценивает возможности и целесообразность использования цифровых технологий в деятельности организации, использует современные цифровые технологии и программные продукты для решения профессиональных задач
ОПК-6.	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
ОПК-6-1.	Знает процессы, методы поиска, сбора, хранения, обработки, представления, распространения информации и способы осуществления таких процессов и методов; современные инструментальные среды, программно-технические платформы и программные средства
ОПК-6.2.	Рационально выбирает современные информационные технологии для решения задач профессиональной деятельности

3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Название раздела/темы	Всего часов	Трудоемкость (час.) по видам учебных занятий			
		Контактная работа			Самостоятельная работа
		Всего	Лекции	семинары	
Тема 1. Введение в большие данные	22	4	2	2	18
Тема 2. Жизненный цикл аналитики данных	26	8	4	4	18
Тема 3. Высокопроизводительные вычисления	26	8	4	4	18
Тема 4. Масштабирование и	26	8	4	4	18

многоуровневое хранение данных					
Тема 5. Визуализация данных и результатов анализа	28	8	4	4	20
Тема 6. Статистические методы анализа данных	26	6	4	2	20
Тема 7. Анализ текста	26	6	2	4	20
Итого	180	48	24	24	132

Тема 1. Введение в большие данные

Большие данные: определение, терминология. Причины появления Big Data. Характеристики Big Data. Источники Big Data. Основные типы данных. Аналитика данных.

Тема 2. Жизненный цикл аналитики данных

Понятие жизненного цикла аналитики данных. Business Intelligence. BI-платформа. ETL – процесс. Средства BI OLAP, OLAP - Многомерный куб. Продвинутая визуализация. Предиктивное моделирование и Data Mining. Инструменты анализа. Power Query. MS Power BI. Pyramid Analytics. Компоненты аналитики MS. SQL server (MDS, SSIS, SSAS). Business Intelligence vs. Data Science.

Тема 3. Высокопроизводительные вычисления

История Hadoop и Map Reduce. Hadoop Distributed File System. Технология Map Reduce. Достоинства модели Map Reduce. Недостатки модели Map Reduce. Архитектура Hadoop. Hadoop MapReduce 1.0. Hadoop Map Reduce 2.0, Варианты использования Hadoop. Экосистема Hadoop. Spark.

Тема 4. Масштабирование и многоуровневое хранение данных

Масштабируемость. Репликация. CAP – теорема. Основы NoSQL. MongoDB.

Тема 5. Визуализация данных и результатов анализа

Типы визуализации. Задачи визуализации. Требования к визуализации. Традиционные виды визуализации. Графики и диаграммы. Инфографика. Презентация и анализ данных. Интерактивный сторителлинг. Дашборды и бизнес аналитика. Визуализация в медицине и науке. Карты и картограммы. Облако тегов. Кластерграмма. Исторический поток. Пространственный поток. Язык R, основные возможности R. Amazon S3. Многокомпонентная загрузка, особенности хранения в S3. Дедупликация данных.

Тема 6. Статистические методы анализа данных

Статистические гипотезы. Статистические критерии. Машинное обучение. Метрический классификатор. Линейный классификатор. ROC – кривая. Кластерный анализ. Алгоритм K-means. Алгоритм C-means. Поиск ассоциативных правил. Свойство антимонотонности.

Тема 7. Анализ текста

Особенности анализа текста, варианты использования. Полнотекстовый поиск. Lucene. Solr. Elasticsearch. Word2Vec.

4. ОЦЕНОЧНЫЕ СРЕДСТВА И ПРИМЕРЫ ЗАДАНИЙ ДЛЯ ОЦЕНКИ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ

4.1 Текущий контроль

Оценка за курс складывается из следующих видов заданий текущего контроля:

Компоненты	Доля в общей оценке
Домашние задания	60%
Решение задач в классе	20%
Экзамен	20%

На курсе используется 10 балльная система оценивания. За каждое задание студент получает от 1 до 10 баллов. Итоговый балл за каждый вид заданий рассчитывается как среднее арифметическое всех полученных баллов за все задания в рамках одного вида (O1, O2, O3). Невыполненное в срок задание оценивается в 0 баллов.

Общая оценка за курс (O) рассчитывается как:

$$O = O1 \times 0,6 + O2 \times 0,2 + O3 \times 0,2.$$

Если по результатам текущего контроля студент получил положительную оценку (не ниже “удовлетворительно”). Оценка за промежуточную аттестацию выставляется автоматически.

Домашние задания

По каждой теме студентам будут предложены задачи, требующие понимания изученного материала

Решение задач в классе

Студентам будут предложены несколько заданий, которые необходимо решить самостоятельно или в группе и представить на занятии.

Экзамен

Во время экзамена студентам будет необходимо решить предложенные преподавателем задачи. Критерии оценивания экзамена представлены в разделе 4.2

Примеры заданий представлены в разделе 4.3

4.2 Промежуточная аттестация

Студентам, набравшим достаточные для удовлетворительной оценки баллы за текущий контроль, оценка за дисциплину выставляется равной оценке за текущий контроль (См. п. 4.1)

Студентам, получившим неудовлетворительную оценку по результатам текущего контроля, необходимо сдать экзамен в письменной форме.

Продолжительность письменного экзамена - 4 академических часа.

Студентам будут предложены два задания. Одно - теоретический вопрос в рамках тематики дисциплины, второй - задание по анализу или визуализации данных.

Критерии оценивания экзамена

Оценка	Критерий
--------	----------

5 Отлично	10	Студент продемонстрировал всесторонние, систематизированные, глубокие знания и умение уверенно применять их на практике при решении конкретных задач, свободное и безупречное обоснование принятых решений
	9	Студент продемонстрировал всесторонние, систематизированные, глубокие знания и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений
4 Хорошо	8	Студент продемонстрировал всесторонние, систематизированные, знания и умение применять их на практике при решении конкретных задач, правильное обоснование принятых решений, но при оформлении работы допущена некоторая небрежность, не влияющее на качество изложения теоретического материала и представление решения практической задачи
	7	Студент твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе на теоретические вопросы некоторую неполноту, которую может устранить с помощью дополнительных вопросов преподавателя
3 Удовлетворительно	6	Студент знает основной материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности, которые может устранить с помощью дополнительных вопросов преподавателя
	5	Студент знает основной материал, по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении практических задач неполноту и неточности, некоторые из которых может устранить только с помощью наводящих вопросов преподавателя
2 Неудовлетворительно	4	Студент продемонстрировал знание отдельных тем, привел правильные формулировки некоторых базовых понятий, в изложении материала нарушена логическая последовательность; практические задачи может решать по предложенным в рамках дисциплины образцам, не демонстрируя их творческой адаптации под конкретную ситуацию
	3	Студент не продемонстрировал знание материала, есть значительные ошибки в формулировках базовых

		понятий, в изложении материала нарушена логическая последовательность; практические задачи решены с ошибками
	1,2	Студент не знает основного содержания тем дисциплины, допускает грубые ошибки в формулировках основных понятий и/или не решил практическую задачу

4.3 Примеры заданий

Примерная тематика задач для домашней работы и решения в классе

1. Корреляции и регрессионный анализ в области больших данных.
2. Логистическая регрессия.
3. Наивный классификатор Байеса.
4. Алгоритм априори.
5. Алгоритм k-means.
6. Визуализация стандартных наборов данных с помощью Tableau.
7. Реализация Hadoop.

Примеры заданий для экзамена

1. Описать группировку и детализацию данных в Tableau, контент-фильтр
2. Дать определение больших данных, подкрепить примерами их использования в бизнесе.
3. Определить основные характеристики BigData.
4. Подробно перечислить основные принципы MapReduce.
5. Описать технологии хранения данных.
6. Привести примеры готовых решений анализа данных, их основные характеристики. Достоинства и недостатки.
7. Описать жизненный цикл анализа данных на конкретном примере решаемой бизнес-задачи.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Литература

Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2024. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536007> (дата обращения: 23.07.2024).

5.2 Электронные образовательные ресурсы

Материалы дисциплины размещены в LMS: <https://l.skolkovo.ru/login/index.php>

5.3 Профессиональные базы данных и информационные справочные системы (при наличии)

нет

6. ЛИЦЕНЗИОННОЕ И СВОБОДНО РАСПРОСТРАНЯЕМОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Операционная система Simple Linux, браузер Yandex браузер, антивирусное ПО Calmantivirus;

Свободно распространяемое ПО, в том числе отечественного производства:

Офисный пакет Libre Office, Okular PDF Reader, 7-Zip Архиватор, GIMP Редактирования фотографий, Inkscape Векторная графика, Blender 3D графика, Kdenlive Видеоредактор, Audacity Аудиоредактор, VLC Медиаплеер, Thunderbird Почтовый клиент, Flameshot Создание скриншотов

7.МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебная аудитория для проведения занятий лекционного типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Учебная аудитория для проведения занятий семинарского типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Аудитория (коворкинг) для самостоятельной работы оснащенная учебной мебелью, ноутбуками.

Материально-техническое обеспечение аудиторий представлено на официальном сайте <https://bbask.ru/sveden/objects/>