

Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования
Московский государственный университет имени М.В. Ломоносова
Механико-математический факультет



УТВЕРЖДАЮ
декан механико-
математического факультета
/А.И. Шафаревич /
«14» октября 2021г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:
Аналитика больших данных

Уровень высшего образования:
магистратура

Направление подготовки / специальность:
02.04.01 "Математика и компьютерные науки" (3++)

Направленность (профиль) ОПОП:
Цифровые технологии и искусственный интеллект

Форма обучения:
очная

Рабочая программа рассмотрена и утверждена
на заседании Ученого совета Механико-математического факультета
(протокол №7, от 14 октября 2021 года)

Москва 2021

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 02.04.01 "Математика и компьютерные науки" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. Место дисциплины (модуля) в структуре ОПОП ВО:
дисциплина относится к вариативной части ОПОП ВО.

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть): отсутствуют.

3. Результаты обучения по дисциплине (модулю):

| Планируемые результаты обучения по дисциплине (модулю) | | |
|--|--|--|
| Содержание и код компетенции. | Индикатор (показатель) достижения компетенции | Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций |
| ОПК-5. Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований | ОПК-5.1. Адаптирует известные научные принципы и методы исследований с целью их практического применения | ОПК-5.1. З-1. Знает фундаментальные научные принципы и методы исследований ОПК-5.1. У-1. Умеет адаптировать с целью практического применения фундаментальные и новые научные принципы и методы исследований |
| ОПК-8. Способен применять методы системного анализа и программное обеспечение для системного моделирования с целью решения задач в сфере исследовательской деятельности | ОПК-8.2. Настраивает, конфигурирует и адаптирует программные средства системного моделирования для постановки и решения задач в сфере исследовательской деятельности | ОПК-8.2. З-1. Знает основные программные средства, используемые для системного моделирования в сфере исследовательской деятельности ОПК-8.2. З-2. Знает принципы работы, системную архитектуру и основные технические характеристики программных средств, используемых для системного моделирования в сфере исследовательской деятельности ОПК-8.2. У-1. Умеет сформулировать задачу и гипотезу исследования с использованием программного кода средств системного моделирования ОПК-8.2. У-2. Умеет конфигурировать и адаптировать типовые программные средства системного анализа и моделирования для |

| | | |
|---|---|--|
| | | решения задач в сфере исследовательской деятельности |
| ПК-2. Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования | ПК-2.1. Выбирает и разрабатывает программные компоненты систем искусственного интеллекта | ПК-2.1. З-1. Знает основные критерии эффективности и качества функционирования системы искусственного интеллекта: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем искусственного интеллекта ПК-2.1. З-2. Знает методы, языки и программные средства разработки программных компонентов систем искусственного интеллекта ПК-2.1. У-1. Умеет выбирать, адаптировать, разрабатывать и интегрировать программные компоненты систем искусственного интеллекта с учетом основных критериев эффективности и качества функционирования |
| 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-6.2. Применяет варианты использования больших данных, определений, словарей и эталонной архитектуры больших данных при руководстве проектами по построению комплексных систем на основе аналитики больших данных в различных отраслях ПК-6.3. Проводит планирование, управление, развертывание, аудит безопасности и защиты персональных данных при работе с большими данными и руководит операционной деятельностью, связанной с безопасностью и | ПК-6.1. З-1. Знает методологию и принципы руководства проектом по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных ПК-6.1. З-2. Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных 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 ПК-6.1. У-2. Умеет сосредотачивать внимание на целях, достижение которых обеспечивает большую отдачу и сильное воздействие |

| | | |
|--|--|--|
| | <p>защитой персональных данных при работе с большими данными</p> | <p>ПК-6.1. У-3. Умеет формировать матрицу приоритетов, включая критерии отбора проектов для реализации</p> <p>ПК-6.2. У-1. Умеет определять риски, связанные с реализацией / развертыванием инициатив / проектов в области аналитики больших данных</p> <p>ПК-6.2. У-2. Умеет описывать каждый риск на различных этапах развертывания аналитики больших данных, его воздействие, реализацию и серьезность</p> <p>ПК-6.2. У-3. Умеет определять цели проектов в области аналитики больших данных в организации / подразделениях / службах</p> <p>ПК-6.2. У-4. Умеет разрабатывать стратегические планы на уровне организации для проектов аналитики больших данных</p> <p>ПК-6.3. З-1. Знает терминологию и последовательность мероприятий по безопасности и защите персональных данных при работе с большими данными</p> <p>ПК-6.3. У-1. Умеет проводить подготовку и планирование действий по верхнеуровневому управлению безопасностью и защитой персональных данных при работе с большими данными</p> <p>ПК-6.3. У-2. Умеет проводить мониторинг, оценку и контроль действий по верхнеуровневому управлению безопасностью и защитой персональных данных при работе с большими данными</p> <p>ПК-6.3. У-3. Умеет определять цели верхнеуровневого управления безопасностью и защитой персональных данных при работе с большими данными</p> |
|--|--|--|

4. Объем дисциплины (модуля) составляет 3 з.е., в том числе 20 академических часов, отведенных на контактную работу обучающихся с преподавателем, 88 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

5.1. Структура дисциплины (модуля) по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

| Наименование разделов и тем дисциплины (модуля), Форма промежуточной аттестации по дисциплине (модулю) | Номинальные трудозатраты обучающегося | | | | |
|---|---------------------------------------|--|--------------------|--------------------------------|-----------------------------------|
| | Всего, ак. ч. | Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, ак. ч. | | | Самостоятельная работа, ак. ч. |
| | | Ауд., ак. ч. | Лекции*, ак. ч. | Семинары*, прак., ак. ч. | |
| Устройство виртуальной машины Java. | 7 | 2 | 2, опрос | | 5 |
| Функциональное программирование (Scala). | 7 | 2 | 2, опрос | | 5 |
| Источники данных и работа с ними. | 7 | 1 | 1, опрос | | 6 |
| Инструменты для написания кода. | 7 | 1 | 1, опрос | | 6 |
| Apache Spark и обработка данных (Scala + Python). | 7 | 1 | 1, опрос | | 6 |
| Экосистема Hadoop. | 7 | 1 | 1, опрос | | 6 |
| Hadoop как хранилище данных. | 7 | 1 | 1, опрос | | 6 |
| Apache Airflow для оркестрации конвейеров. | 7 | 1 | 1, опрос | | 6 |
| Облачные хранилища. | 8 | 1 | 1, опрос | | 7 |
| Распределённые системы. | 8 | 1 | 1, опрос | | 7 |

| | | | | | |
|--|---|---|----------|--|----|
| Архитектура обработки больших данных. | 8 | 2 | 2, опрос | | 6 |
| Алгоритмы параллельной обработки данных. | 8 | 2 | 2, опрос | | 6 |
| Программные платформы для систем больших данных. | 8 | 2 | 2, опрос | | 6 |
| Оборудование для простроения ЦОД | 8 | 2 | 2, опрос | | 6 |
| Промежуточная аттестация - устный экзамен. | 4 | | | | 4 |
| Итого | | | 20 | | 88 |

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания (в отсутствие утвержденных соответствующих локальных нормативных актов на факультете)

Примерные темы для опросов:

1. Устройство виртуальной машины Java.
2. Функциональное программирование (Scala).
3. Источники данных и работа с ними.
4. Инструменты для написания кода.
5. Apache Spark и обработка данных (Scala + Python).
6. Экосистема Hadoop.
7. Hadoop как хранилище данных.
8. Apache Airflow для оркестрации конвейеров.
9. Облачные хранилища.
10. Распределённые системы.
11. Архитектура обработки больших данных.
12. Алгоритмы параллельной обработки данных.
13. Программные платформы для систем больших данных.
14. Оборудование для простроения ЦОД.

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине, критерии и шкалы оценивания Примерные вопросы для проведения экзамена:

Примерные вопросы для проведения экзамена:

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
2. Виды массивов данных.
3. Базовые принципы обработки больших данных.
4. Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
5. Технологии Business Intelligence и реляционные системы управления базами данных.
6. Прогнозирование и предвидение: общее и особенное.
7. Виды прогнозов.
8. Опишите методики анализа больших данных.
9. Процесс аналитики анализа больших данных.
10. Дайте характеристику Big Data на мировом рынке.
11. Охарактеризуйте Big Data в России.
12. Определите понятие Data Mining.
13. Вопросы безопасности больших данных.
14. В чем состоит когнитивный анализ данных.
15. Какие модели данных вы знаете?
16. Основные описательные статистики.
17. Определите различия между параметрическими, непараметрическими и номинальными методами.
18. Опишите основную идею корреляционного анализа.
19. Регрессионный анализ.
20. Основная идея дисперсионного анализа.
21. Сущность кластерного анализа.
22. Дискриминантный анализ: модель и общая процедура выполнения.
23. Цели факторного анализа.
24. Программные средства анализа данных: Statistica, SPSS, Excel; их преимущества и недостатки.
25. Языки для Big Data.
26. Фреймворки для Big Data.
27. Базы данных для Big Data.
28. Аналитические платформы для Big Data.

Порядок организации оценивания результатов обучения в университете регламентируется отдельным локальным нормативным актом – «Система оценивания результатов обучения в Томском политехническом университете (Система оценивания)» (в действующей редакции). Используется балльно-рейтинговая система оценивания результатов обучения. Итоговая оценка (традиционная и литерная) по видам учебной деятельности (изучение дисциплин, УИРС, НИРС, курсовое проектирование, практики) определяется суммой баллов по результатам текущего контроля и промежуточной аттестации (итоговая рейтинговая оценка - максимум 100 баллов).

Распределение основных и дополнительных баллов за оценочные мероприятия текущего контроля и промежуточной аттестации устанавливается календарным рейтингом-планом дисциплины.

Рекомендуемая шкала для отдельных оценочных мероприятий входного и текущего контроля

| % выполнения задания | Соответствие традиционной оценке | Определение оценки |
|-----------------------------|---|--|
| 90%÷100% | «Отлично» | Отличное понимание предмета, всесторонние знания, отличные умения и владение опытом практической деятельности, необходимые результаты обучения сформированы, их качество оценено количеством баллов, близким к максимальному |
| 70% - 89% | «Хорошо» | Достаточно полное понимание предмета, хорошие знания, умения и опыт практической деятельности, необходимые результаты обучения сформированы, качество ни одного из них не оценено минимальным количеством баллов |
| 55% - 69% | «Удовл.» | Приемлемое понимание предмета, удовлетворительные знания, умения и опыт практической деятельности, необходимые результаты обучения сформированы, качество некоторых из них оценено минимальным количеством баллов |
| 0% - 54% | «Неудовл.» | Результаты обучения не соответствуют минимально достаточным требованиям |

Шкала для оценочных мероприятий Экзамена

| % выполнения заданий Экзамена | Экзамен, балл | Соответствие традиционной оценке | Определение оценки |
|--------------------------------------|----------------------|---|--|
| 90%÷100% | 18 ÷ 20 | «Отлично» | Отличное понимание предмета, всесторонние знания, отличные умения и владение опытом практической деятельности, необходимые результаты обучения сформированы, их качество оценено количеством баллов, близким к максимальному |
| 70% - 89% | 14 ÷ 17 | «Хорошо» | Достаточно полное понимание предмета, хорошие знания, умения и опыт практической деятельности, необходимые результаты обучения сформированы, качество ни одного из них не оценено минимальным количеством баллов |
| 55% - 69% | 11 ÷ 13 | «Удовл.» | Приемлемое понимание предмета, удовлетворительные знания, умения и опыт практической деятельности, необходимые результаты обучения сформированы, качество некоторых из них оценено минимальным количеством баллов |
| 0% - 54% | 0 ÷ 10 | «Неудовл.» | Результаты обучения не соответствуют минимально достаточным требованиям |

Методические указания по процедуре оценивания

| № | Оценочные мероприятия | Процедура проведения оценочного мероприятия и необходимые методические указания | | | | |
|-------------|---|--|-------------|-----------------|--|--|
| 1. | Опрос (макс 1 балл) | <ul style="list-style-type: none"> ● Письменный опрос проводится по пройденному материалу в течение первых 5-10 минут занятия ● Опрос содержит 5 вопросов ● Каждый вопрос оценивается в 0,2 балла. ● Опрос считается успешно выполненным при получении более 0,5 балла за все вопросы. <p>Оценивание проводит преподаватель по следующим критериям: 0,2 – студент полно и правильно отвечает на вопрос; 0,15 – студент дал неполный ответ на вопрос, но не допускает ошибок; 0,1 – студент допускает отдельные существенные ошибки, но понимает суть вопроса и основные закономерности; 0,05 – студент излагает материал со значительными ошибками, демонстрирует слабое понимание сути вопроса; 0 – нет ответа.</p> | | | | |
| 2. | Тестирование (макс 4 балла) | <ul style="list-style-type: none"> ● Письменное тестирование проводится после изучения теоретического материала раздела и обсуждения на лабораторных занятиях по каждой теме. ● Задание содержит 5 вопросов в тестовой форме ● Вопрос на выбор из предложенных вариантов правильной информации оценивается в 0,8 балла. ● Тест считается успешно выполненным при получении более 2 баллов за все задание. <p>Оценивание проводит преподаватель по следующим критериям: 0,8 – выбран правильный ответ (ответы); 0,5 – выбраны большинство правильных ответов; 0 – выбраны правильные и неправильные ответы.</p> | | | | |
| 3. | Защита лабораторной работы (макс 15 баллов) | <ul style="list-style-type: none"> ● Защита лабораторных работ проводится на 4, 8, 12 и 16 неделе, соответственно. ● Отчет по лабораторной работе содержит полную информацию о результатах работы магистранта в ходе лабораторных работ в соответствии с требованиями, изложенными в методических указаниях и заданным вариантом. ● При защите лабораторной работы магистрант обосновывает структуру СМО, доказывает полноту проведенных экспериментов, проводит анализ на основе полученных характеристик модели. ● Отчет по лабораторной работе считается успешно защищенным при получении более 7,5 баллов. <p><u>Оценивание проводит преподаватель по следующим критериям:</u></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%; text-align: center;">Вид вопроса</th> <th style="text-align: center;">Критерии оценки</th> </tr> </thead> <tbody> <tr> <td style="height: 20px;"> </td> <td> </td> </tr> </tbody> </table> | Вид вопроса | Критерии оценки | | |
| Вид вопроса | Критерии оценки | | | | | |
| | | | | | | |

| | | | | | | | | | | | | | | |
|---|--|---|--|---|---|--|---|--|---|--|--------------------|--|---|---|
| | | <table border="1"> <tr> <td>Математическая логика, теория алгоритмов и теория сложности</td> <td>5 б. – Модель корректная, дополнена необходимыми блоками.</td> <td>3 б. – Модель корректная, модификация проведена не полностью.</td> <td>1 б. – Модель работает не во всех случаях корректно, модификация не проведена.</td> </tr> <tr> <td>Планирование экспериментальных исследований</td> <td>5 б. – Запланирован и проведен полный цикл экспериментальных исследований.</td> <td>3 б. – Запланирован и проведен неполный цикл экспериментальных исследований, не позволяющий провести корректный анализ.</td> <td>1 б.– Проведенные исследования не позволяют сделать корректные выводы, отсутствуют необходимые эксперименты, проведены избыточные.</td> </tr> <tr> <td>Анализ результатов</td> <td>5 б.– Проведен корректный и полный анализ. Все выводы подкреплены достаточным графическим и табличным результатом.</td> <td>3 б. – Проведен неполный анализ, табличных и графических материалов недостаточно.</td> <td>1 б.– При анализе результатов сделаны некорректные выводы, графические и табличные результаты с ошибками.</td> </tr> </table> | Математическая логика, теория алгоритмов и теория сложности | 5 б. – Модель корректная, дополнена необходимыми блоками. | 3 б. – Модель корректная, модификация проведена не полностью. | 1 б. – Модель работает не во всех случаях корректно, модификация не проведена. | Планирование экспериментальных исследований | 5 б. – Запланирован и проведен полный цикл экспериментальных исследований. | 3 б. – Запланирован и проведен неполный цикл экспериментальных исследований, не позволяющий провести корректный анализ. | 1 б.– Проведенные исследования не позволяют сделать корректные выводы, отсутствуют необходимые эксперименты, проведены избыточные. | Анализ результатов | 5 б.– Проведен корректный и полный анализ. Все выводы подкреплены достаточным графическим и табличным результатом. | 3 б. – Проведен неполный анализ, табличных и графических материалов недостаточно. | 1 б.– При анализе результатов сделаны некорректные выводы, графические и табличные результаты с ошибками. |
| Математическая логика, теория алгоритмов и теория сложности | 5 б. – Модель корректная, дополнена необходимыми блоками. | 3 б. – Модель корректная, модификация проведена не полностью. | 1 б. – Модель работает не во всех случаях корректно, модификация не проведена. | | | | | | | | | | | |
| Планирование экспериментальных исследований | 5 б. – Запланирован и проведен полный цикл экспериментальных исследований. | 3 б. – Запланирован и проведен неполный цикл экспериментальных исследований, не позволяющий провести корректный анализ. | 1 б.– Проведенные исследования не позволяют сделать корректные выводы, отсутствуют необходимые эксперименты, проведены избыточные. | | | | | | | | | | | |
| Анализ результатов | 5 б.– Проведен корректный и полный анализ. Все выводы подкреплены достаточным графическим и табличным результатом. | 3 б. – Проведен неполный анализ, табличных и графических материалов недостаточно. | 1 б.– При анализе результатов сделаны некорректные выводы, графические и табличные результаты с ошибками. | | | | | | | | | | | |
| | | При несвоевременной сдаче лабораторной работы оценка снижается на 1 балл за каждую просроченную неделю. | | | | | | | | | | | | |
| 4. | Экзамен | <p>В рамках изучаемых разделов дисциплины осуществляется текущее оценивание степени освоения студентами изученного материала. Проверка освоения лекционного материала проводится путем оценки результатов выполнения лабораторных работ.</p> <p>Допуск по итогу текущего контроля рассчитывается на основе суммы баллов, набранных за все виды оценочных мероприятий. Для допуска к Экзамену студенту необходимо набрать 55 баллов и более по всем видам запланированных оценочных мероприятий, при этом все виды запланированных оценочных мероприятий должны быть выполнены и зачтены преподавателем.</p> <p>Экзамен проводится в традиционной форме путём раздачи билетов, самостоятельной подготовки студентами ответов на вопросы билета, последующей беседы преподавателя со студентом.</p> <p>Экзаменационный билет состоит из 4 вопросов.</p> <p>Каждый вопрос оценивается преподавателем исходя из максимального балла – 5 баллов. Максимальный балл за Экзамен 20 баллов.</p> | | | | | | | | | | | | |

| | |
|--|--|
| | Итоговая отметка за семестр рассчитывается на основе полученной суммы баллов в результате текущего контроля, и баллов, набранных при заключительном контроле знаний на Экзамене. |
|--|--|

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

1. Анналин, Ын., Теоретический минимум по Big Data. СПб.: Издательство: Питер, 2019, 208 с. ISBN: 978-5-4461-1040-7.
2. Базы данных в высокопроизводительных информационных системах : учебное пособие / авт.-сост. Е.И. Николаев ; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь : СКФУ, 2016 - 163 с. : ил. - Библиогр.: с.161
3. Онлайн курс Big Data Fundamentals via PySpark, автор Upendra Kumar Devisetty Science Analyst at CyVerse <https://www.datacamp.com/courses/big-data-fundamentals-via-pyspark>

Дополнительная литература

1. Minelli M., Chambers M., Dhiraj A. Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses. John Wiley & Sons, 2012.
2. Сенько, А. Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure. СПб: Издательство: Питер, 2018, 448 с. ISBN: 978-5-4461-0578-6.
3. Чак Лэм, Nadoor в действии - ДМК Пресс, 2012

7.2. Перечень профессиональных баз данных и информационных справочных систем

1. Du D.-Z., Ko K.-I. Theory of computational complexity. – John Wiley & Sons, 2000.
2. Sipser M. Introduction to the theory of computation. – Thomson Course Technology, 2nd ed., 2006.
3. Лавров, Игорь Андреевич. Математическая логика: учебное пособие для студентов высших учебных заведений, обучающихся по техническим и естественнонаучным специальностям / И. А. Лавров; под ред. Л. Л. Максимовой. Москва: Академия, 2006, 39, [1] с.: ил; 22 см. (Университетский учебник. Прикладная математика и информатика. ISBN 5-7695-2735-8.
4. Мендельсон, Э. Введение в математическую логику / Э. Мендельсон; под ред. С.И. Адян; пер. с англ. Ф.А. Кабакова. - Москва: Наука, 1971 - 320 с.: ил.; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=458257>.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://window.edu.ru/> Электронная библиотека учебно-методической литературы для общего и профессионального образования;
2. <http://elibrary.ru/> Научная электронная библиотека;
- <http://www.scholar.ru/> Научные статьи, диссертации и авторефераты из электронных научных библиотек.

8.2. Описание материально-технического обеспечения.

| № | Наименование | Назначение |
|----|---|---|
| 1. | Презентационное оборудование (мультимедиапроектор, экран, компьютер для управления) | Для проведения лекционных занятий |
| 2. | Компьютерный класс (с выходом в Internet) | Для организации самостоятельной работы и организации практических занятий обучающихся |

9. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

10. Разработчик (разработчики) программы:

Михаил Александрович Королёв, Ведущий Data Engineer в компании «Альфастрахование»

Кирилл Анатольевич Власов, инженер компании «Цифра».