



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»



УТВЕРЖДАЮ
Директор ИЕиС
И.Ю. Мезин

19.02.2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

БОЛЬШИЕ И ОТКРЫТЫЕ ДАННЫЕ

Направление подготовки (специальность)
01.03.02 Прикладная математика и информатика

Направленность (профиль/специализация) программы
Большие и открытые данные

Уровень высшего образования - бакалавриат

Форма обучения
очная

Институт/ факультет	Институт естествознания и стандартизации
Кафедра	Прикладной математики и информатики
Курс	2
Семестр	4

Магнитогорск
2024 год

Рабочая программа составлена на основе ФГОС ВО - бакалавриат по направлению подготовки 01.03.02 Прикладная математика и информатика (приказ Минобрнауки России от 10.01.2018 г. № 9)

Рабочая программа рассмотрена и одобрена на заседании кафедры Прикладной математики и информатики
09.02.2024, протокол № 6

Зав. кафедрой  Ю.А. Извеков

Рабочая программа одобрена методической комиссией ИЕиС
19.02.2024 г. протокол № 5

Председатель  И.Ю. Мезин

Рабочая программа составлена:
доцент кафедры ПМИИ, канд. пед. наук

 Т.П. Злыднева

Рецензент:

зав. кафедрой Физики, канд. физ.-мат. наук



Д.М. Долгушин

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2025 - 2026 учебном году на заседании кафедры Прикладной математики и информатики

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Ю.А. Извеков

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2026 - 2027 учебном году на заседании кафедры Прикладной математики и информатики

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Ю.А. Извеков

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2027 - 2028 учебном году на заседании кафедры Прикладной математики и информатики

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Ю.А. Извеков

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2028 - 2029 учебном году на заседании кафедры Прикладной математики и информатики

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Ю.А. Извеков

1 Цели освоения дисциплины (модуля)

Целями освоения дисциплины «Большие и открытые данные» являются: изучение математических методов и подходов, используемых в программных системах обработки и анализа больших данных компании SAS - мирового лидера в области разработки и внедрения IT решений и услуг в бизнес-аналитике, в том числе в финансовой и банковской сфере; развитие профессиональных навыков учащихся за счет получения практического опыта работы с IT решениями компании SAS в части обработки и анализа больших данных.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Большие и открытые» данные входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Информатика

Программная инженерия

Практикум на ЭВМ

Теория языков программирования

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Базы данных

Выполнение и защита выпускной квалификационной работы

Производственная - научно-исследовательская работа

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Большие и открытые данные» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-1	Способен анализировать большие данные и проводить исследования с их технологиями
ПК-1.1	Выбирает методы и инструментальные средства для проведения аналитических работ с большими данными
ПК-1.2	Разрабатывает и оценивает модели больших данных
ПК-1.3	Организует контроль эффективности работы и предлагает решения руководителю (заказчику)

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 зачетных единиц 108 акад. часов, в том числе:

- контактная работа – 68,95 акад. часов;
- аудиторная – 68 акад. часов;
- внеаудиторная – 0,95 акад. часов;
- самостоятельная работа – 39,05 акад. часов;
- в форме практической подготовки – 0 акад. час;

Форма аттестации - зачет

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Технологии хранения и обработки Больших данных								
1.1 Введение в Большие данные	4	2		2	4	Изучение учебной и научной литературы.	Проверка изучения основной и дополнительной литературы.	
1.2 Модель вычислений MapReduce		2		2	2	Подготовка к лабораторному занятию.	Отчет о выполнении лабораторных работ. Опрос, обсуждение.	
1.3 Реализации алгоритмов на MapReduce. Основы Hive.		2		4	1	Решение задач.	Проверка изучения основной и дополнительной литературы.	
Итого по разделу		6		8	7			
2. Программирование обработки и загрузки Больших данных в SAS								
2.1 Основы языка SAS Base	4	2		2	4	Подготовка к лабораторному занятию.	Опрос, обсуждение. Отчет о выполнении лабораторных работ.	
2.2 Макропроцессор SAS. SAS SQL				2	1	Подготовка к лабораторному занятию.	Отчет о выполнении лабораторных работ.	
2.3 Обзор SAS Data Loader for Hadoop: автоматизированная загрузка и обработка больших данных		2		4	1	Изучение учебной и научной литературы.	Контрольная работа. Отчет о выполнении лабораторных работ.	

2.4 Взаимодействие SAS и Hadoop			2	2	Подготовка к лабораторному занятию.	Отчет о выполнении лабораторных работ.	
2.5 Основы программирования на SAS Data Step 2 (DS2)		2	4	2	Подготовка к лабораторному занятию.	Опрос, обсуждение. Отчет о выполнении лабораторных работ.	
Итого по разделу		6	14	10			
3. Аналитика в больших данных							
3.1 Обзор задач и методов машинного обучения и интеллектуального анализа больших данных	4		2	4	Подготовка к лабораторному занятию.	Опрос, обсуждение.	
3.2 Интерактивное исследование данных с помощью SAS VA/VS			4	1	Изучение учебной и научной литературы.	Опрос, обсуждение.	
3.3 Построение моделей с помощью IMSTAT в SAS LASR сервере			4		Изучение учебной и научной литературы.	Опрос, обсуждение.	
3.4 SAS Enterprise Miner			2	2	Подготовка к лабораторному занятию.	Отчет о выполнении лабораторных работ.	
3.5 Операционализация моделей		2	2		Подготовка к лабораторному занятию.	Отчет о выполнении лабораторных работ.	
Итого по разделу		2	14	7			
4. Аналитическая обработка сложноструктурированных больших данных							
4.1 Обработка текстов в SAS Text Miner	4		4	2	Решение задач.	Проверка изучения основной и дополнительной литературы.	
4.2 Анализ временных рядов (SAS Forecast Studio)		2	4	2	Изучение учебной и научной литературы.	Контрольная работа. Отчет о выполнении лабораторных работ.	
4.3 Анализ взаимосвязей с помощью SAS SNA		1	3	5	Изучение учебной и научной литературы.	Контрольная работа.	
4.4 Решение оптимизационных задач в SAS/OR			4	6,05	Изучение учебной и научной литературы.	Контрольная работа. Отчет о выполнении лабораторных работ.	
Итого по разделу		3	15	15,05			
Итого за семестр		17	51	39,05		зачёт	
Итого по дисциплине		17	51	39,05		зачет	

5 Образовательные технологии

1. Традиционные технологии обучения, предполагающие передачу информации в готовом виде, формирование учебных умений по образцу: лекция-изложение, лекция-объяснение, практические (семинарские) занятия.

Использование традиционных технологий обеспечивает ориентирование студента в потоке информации, связанной с различными подходами к определению сущности, содержания, методов, форм развития и саморазвития личности; самоопределение в выборе оптимального пути и способов личностно-профессионального развития; систематизацию знаний, полученных студентами в процессе аудиторной и самостоятельной работы. Лекционные занятия проводятся с использованием мультимедийных средств. Лабораторные занятия обеспечивают развитие и закрепление умений и навыков определения целей и задач саморазвития, а также принятия наиболее эффективных решений по их реализации. Практические занятия проводятся в компьютерных классах вычислительного центра ФГБОУ ВО «МГТУ».

В ходе проведения лекционных занятий предусматривается: использование электронного демонстрационного материала по темам, требующим иллюстрации работы программных продуктов: MS Word, MS Excel.

В ходе проведения лабораторных занятий предусматривается использование средств вычислительной техники при выполнении индивидуальных заданий и тестирования.

2. Интерактивные формы обучения, предполагающие организацию обучения как продуктивной творческой деятельности в режиме взаимодействия студентов друг с другом и с преподавателем

Использование интерактивных образовательных технологий способствует повышению интереса и мотивации учащихся, активизации мыслительной деятельности и творческого потенциала студентов, делает более эффективным усвоение материала, позволяет индивидуализировать обучение и ввести экстренную коррекцию знаний.

В рамках дисциплины «Большие и открытые данные» предусматривается 14 часов аудиторных занятий, проводимых в интерактивной.

При проведении лабораторных занятий используются групповая работа, технология коллективной творческой деятельности, технология сотрудничества, обсуждение проблемы в форме дискуссии, дебаты, круглый стол. Данные технологии обеспечивают высокий уровень усвоения студентами знаний, эффективное и успешное овладение умениями и навыками в предметной области, формируют познавательную потребность и необходимость дальнейшего самообразования, позволяют активизировать исследовательскую деятельность, обеспечивают эффективный контроль усвоения знаний.

3. Возможности образовательного портала ФГБОУ ВО «МГТУ» для предоставления студентам графика самостоятельной работы, расписания консультаций, заданий для самостоятельного выполнения и рекомендуемых тем для самостоятельного изучения.

Методика, предлагаемая для изучения курса «Большие и открытые данные» ориентирована на лекции проблемно-информационного характера, семинарские занятия исследовательского типа и подготовку рефератов.

Используемые образовательные технологии позволяют активно применять в учебном процессе интерактивные формы проведения занятий (компьютерная симуляция, разбор конкретных ситуаций), что способствует формированию и развитию профессиональных навыков обучающихся. Применяемые в процессе изучения дисциплины поисковый и исследовательский методы в полной мере соответствуют требованиям ФГОС по реализации компетентностного подхода.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)**а) Основная литература:**

Огнева, М. В. Программирование на языке C++: практический курс : учебное пособие для вузов / М. В. Огнева, Е. В. Кудрина. — Москва : Издательство Юрайт, 2021. — 335 с. — (Высшее образование). — ISBN 978-5-534-05123-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/473054>

б) Дополнительная литература:

Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2021. — 161 с. — (Высшее образование). — ISBN 978-5-534-10971-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/472985>

в) Методические указания:

Маркин, А. В. Программирование на SQL в 2 ч. Часть 1 : практикум для вузов / А. В. Маркин. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2021. — 403 с. — (Высшее образование). — ISBN 978-5-534-12256-5. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/471148>

г) Программное обеспечение и Интернет-ресурсы:**Программное обеспечение**

Наименование ПО	№ договора	Срок действия лицензии
MS Office 2007 Professional	№ 135 от 17.09.2007	бессрочно
7Zip	свободно	бессрочно
Maple 14 Classroom License	К-113-11 от 11.04.2011	бессрочно
MathCAD v.15 Education University	Д-1662-13 от 22.11.2013	бессрочно
WordPress	свободно	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Электронные ресурсы библиотеки МГТУ им. Г.И. Носова	https://magtu.informsystema.ru/Marc.html?locale=ru
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Национальная информационно-аналитическая система – Российский индекс научного	URL: https://elibrary.ru/project_risc.asp

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Лекционная аудитория. Оснащение аудитории: доска, мультимедийные средства хранения, передачи и представления информации

Учебные аудитории для проведения лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации: компьютерный класс. Оснащение аудитории: персональные компьютеры с пакетом MS Office, Maple 14 Classroom License 10-29 Users (per User) Academic, MathLab, Mathcad Education - University Edition (200 pack) и выходом в Интернет.

Помещения для самостоятельной работы обучающихся. Оснащение аудитории: персональные компьютеры с пакетом MS Office, выходом в Интернет и с доступом в электронную информационно-образовательную среду университета.

Помещение для хранения и профилактического обслуживания учебного оборудования. Оснащение аудитории: стеллажи для хранения учебно-наглядных пособий и учебно-методической документации.

Учебно-методическое обеспечение самостоятельной работы

По дисциплине «Большие и открытые данные» предусмотрена аудиторная и внеаудиторная самостоятельная работа обучающихся.

Аудиторная самостоятельная работа студентов предполагает выполнение заданий лабораторных работ.

Примерные аудиторные задания:

1. Терминология, история появления.
2. Технические сложности работы с большими данными.
3. Распределенная файловая система HDFS.
4. Базовая идея модели MapReduce, пример использования MapReduce.
5. Модель вычислений MapReduce.
6. Реализация MapReduce в Hadoop.
7. Основы Java API.
8. Планирование вычислений. YARN.
9. Примеры реализации алгоритмов на MapReduce (включая умножение матриц, операции реляционной алгебры). Колоночные форматы хранения (на примере ORC).
10. Основы Hive.
11. Изучение основ программирования на языке SAS: структуры языка, запуска и отладки программы, простейших аналитических процедур и приёмов для трансформации данных.

Примерные индивидуальные домашние задания (ИДЗ):

1. Дан набор данных заданной структуры и программа SAS Data step, производящая определенную обработку и вычисления с использованием данного набора. Перепишите эту программу на SAS DS2 с использованием параллельных нитей и созданием пользовательского пакета, чтобы результат обработки сохранился тем же, но код мог выполняться в параллельной среде.
2. Дан набор заданной структуры, постройте модель прогнозирования от-клика с использованием процедуры impstat с алгоритмом random forest с заданным числом деревьев. Примените полученную модель к тестовому набору данных той же структуры, визуализируйте полученный график Lift. Постройте на том же наборе модель с использованием высокопроизводительной версии метода GLM. Примените к тестовому набору. Сравните результаты GLM и Random Forest по AUC.

Примерные аудиторные контрольные работы (АКР)

Дан текстовый корпус документов, лежащих в указанной директории. Создайте в SAS Text Miner проект, который: выберет файлы с расширением pdf; осуществит парсинг набора с определением частей речи и сохранением в признаковом пространстве только существительных и глаголов; осуществит фильтрацию документов и признаков с использованием заданной схемой определения весов лексем (например, на основе tf-idf); выделит заданное количество ключевых тематик по методу SVD. В ответе укажите топ 5 ключевых слов во второй выявленной тематике. Какой документ имеет наибольший вес в этой тематике?

Вопросы для итоговой оценки качества освоения курса:

1. Hadoop как технология хранения и обработки больших данных.
2. Способы загрузки данных в Hadoop. Базовые операции с данными.
3. Профилирование, дедупликация. Выполнение процессов Data Quality внутри Hadoop.
4. Интеграция с SAS In-Memory Analytics Server.
5. Обзор взаимодействия SAS и HADOOP.
6. Интерфейс SAS ACCESS в HADOOP (HIVE).
7. Выполнение выражений Pig, HDFS, Map Reduce че-рез PROC HADOOP.
8. Интерфейс SAS ACCESS в HADOOP (Impala)

Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК – 1: Способен анализировать большие данные и проводить исследования с их технологиями		
ПК - 1.1	Выбирает методы и инструментальные средства для проведения аналитических работ с большими данными	<p>Перечень теоретических вопросов:</p> <ol style="list-style-type: none"> 1. Аналитические методы, применимые к большим объемам данных. 2. Многомерные связи, ассоциации, корреляции. Непрерывность поступающих обновлений как характерная черта анализа больших данных. 3. Примеры обработки неструктурированных данных. 4. Понятия неоднозначности и недостоверности данных. <p>Примерные практические задания:</p> <ol style="list-style-type: none"> 1. Обзор взаимодействия SAS и HADOOP. 1. Интерфейс SAS ACCESS в HADOOP (HIVE). 2. Выполнение выражений Pig, HDFS, Map Reduce через PROC HADOOP. 3. Интерфейс SAS ACCESS в HADOOP (Impala) <p>Индивидуальное домашнее задание: Программирование на языке DS2.</p>
ПК-1.2	Разрабатывает и оценивает модели больших данных	<p>Перечень теоретических вопросов</p> <ol style="list-style-type: none"> 1. Базовый функционал SAS EM - определение простого проекта, загрузка и изменение данных. Регрессионные модели и деревья решений. 2. Поиск стандартных путей по истории процессов, частотный анализ и выявление отклонений. 3. Диагностика и оценка качества моделей. Примеры: кредитные модели и оценка рисков. <p>Примерные практические задания</p> <ol style="list-style-type: none"> 1. Примеры визуализаций базового статистического анализа, доступных в SAS VA/VS. 2. Агрегирование и частотный анализ. 3. Диаграммы переходов в системах с процессами. Кластеризация и факторный анализ. 4. Базовая диагностика моделей. 5. Пример проекта - добавление данных, дизайн визуализаций. <p>Индивидуальное домашнее задание:</p> <ol style="list-style-type: none"> 1. Обзор аналитического сервера inmemory аналитики 2. SAS LASR. Загрузка данных на LASR сервер. Исследование данных. Подготовка данных. Обзор алгоритмов машинного обучения в IMSTAT.

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК-1.3	<p>Организует контроль эффективности работы и предлагает решения руководителю (заказчику).</p>	<p>Перечень теоретических вопросов</p> <ol style="list-style-type: none"> 1. Понятие операционализации моделей. 2. Постановка моделей на регламентное переобучение и применение. 3. Мониторинг качества моделей. Создание шаблонов построения моделей. 4. Автоматическое построение моделей по сегментам данных. 5. Экспорт и применение моделей в СУБД. Применение моделей в инструментах аналитики в реальном времени (SAS Decision Manager, SAS Event Stream Processing). <p>Индивидуальное задание: Формулировка и решение задач линейного, нелинейного и целочисленного программирования с помощью процедуры OPTMODEL.</p> <p>Примерные индивидуальные задания: Процедура PROC OPTGRAPH: разбиение графа, расчет метрик центральности и другие алгоритмы теории графов, анализа сетей и оптимизации.</p>

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Большие и открытые данные» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень форсированности умений и владений, проводится в форме зачета.

Показатели и критерии оценивания:

- на оценку **«отлично»** (5 баллов) – обучающийся демонстрирует высокий уровень сформированности компетенций, всестороннее, систематическое и глубокое знание учебного материала, свободно выполняет практические задания, свободно оперирует знаниями, умениями, применяет их в ситуациях повышенной сложности.
- на оценку **«хорошо»** (4 балла) – обучающийся демонстрирует средний уровень сформированности компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
- на оценку **«удовлетворительно»** (3 балла) – обучающийся демонстрирует пороговый уровень сформированности компетенций: в ходе контрольных мероприятий допускаются ошибки, проявляется отсутствие отдельных знаний, умений, навыков, обучающийся испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
- на оценку **«неудовлетворительно»** (2 балла) – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.
- на оценку **«неудовлетворительно»** (1 балл) – обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач.

Для получения зачета необходимо набрать три балла.