

МИНОБРНАУКИ РОССИИ

**Федеральное государственное бюджетное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Кафедра прикладной математики и информатики

**Авторы-составители: Перескокова Ольга Ивановна
Русаков Сергей Владимирович**

Рабочая программа дисциплины

**КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ БОЛЬШИХ МАССИВОВ
ДАНЫХ**

Код УМК 92197

Утверждено
Протокол №9
от «21» мая 2020 г.

Пермь, 2020

1. Наименование дисциплины

Компьютерные технологии обработки больших массивов данных

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в базовую часть Блока « М.1 » образовательной программы по направлениям подготовки (специальностям):

Направление: **01.04.02** Прикладная математика и информатика
направленность Информационно-аналитические системы в прогнозировании и управлении социально-экономическим развитием

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Компьютерные технологии обработки больших массивов данных** у обучающегося должны быть сформированы следующие компетенции:

01.04.02 Прикладная математика и информатика (направленность : Информационно-аналитические системы в прогнозировании и управлении социально-экономическим развитием)

ОПК.2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач

Индикаторы

ОПК.2.3 Реализует математический метод на языке программирования высокого уровня и/или с помощью специализированных пакетов программ

ОПК.4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности

Индикаторы

ОПК.4.1 Комбинирует и адаптирует современные информационно-коммуникационные технологии для реализации решения математических задач

4. Объем и содержание дисциплины

Направления подготовки	01.04.02 Прикладная математика и информатика (направленность: Информационно-аналитические системы в прогнозировании и управлении социально-экономическим развитием)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	1
Объем дисциплины (з.е.)	4
Объем дисциплины (ак.час.)	144
Контактная работа с преподавателем (ак.час.), в том числе:	48
Проведение лекционных занятий	24
Проведение практических занятий, семинаров	24
Самостоятельная работа (ак.час.)	96
Формы текущего контроля	Защищаемое контрольное мероприятие (1) Итоговое контрольное мероприятие (1) Письменное контрольное мероприятие (1)
Формы промежуточной аттестации	Экзамен (1 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Основы языка программирования Python

Изучаются основные типы данных (str, int, float, bool, None), структуры данных языка Python (списки, кортежи, словари, множества) и приемы эффективной работы с ними.

Знакомство с возможностями библиотек NumPy, SciPy, Pandas, Matplotlib. и их использование для выполнения первичной обработки и анализа данных.

Введение в машинное обучение. Обучение с учителем.

Изучаются возможности библиотеки sclearn. Рассматриваются возможность генерации на наборов данных с заданными свойствами и работа с "игрушечными" наборами данных. Рассматриваются задачи линейной регрессии и задачи классификации. Изучаются способы построения решающих деревьев и лесов решающих деревьев. Изучаются способы борьбы с переобучением моделей.

Практическое применение технологий обработки данных.

Выполняется исследовательская работа, в которой для набора данных проводится первичная обработка и чистка данных. Исследуются зависимости данных. Выбираются модели машинного обучения для решения поставленных задач. Выполняется подбор параметров по сетке для выбранных моделей. Оценивается качество полученных результатов.

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторские занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Пальмов, С. В. Интеллектуальный анализ данных : учебное пособие / С. В. Пальмов. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/75376.html>
2. Сузи, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 350 с. — ISBN 978-5-4497-0705-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/97589.html>
3. Федин, Ф. О. Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 308 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/26445>
4. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2019. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/432851>
5. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие / Ф. О. Федин, Ф. Ф. Федин. — Москва : Московский городской педагогический университет, 2012. — 204 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/26444>
6. Сузи, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 350 с. — ISBN 978-5-4497-0705-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/97589.html>

Дополнительная:

1. Буйначев, С. К. Основы программирования на языке Python : учебное пособие / С. К. Буйначев, Н. Ю. Боклаг ; под редакцией Ю. В. Песин. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2014. — 92 с. — ISBN 978-5-7996-1198-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/66183.html>
2. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2019. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. <https://www.urait.ru/bcode/432178>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<http://www.psu.ru/elektronnye-resursy-dlya-psu> Электронные ресурсы для ПГНИУ

<http://window.edu.ru/> Единое окно доступа к образовательным ресурсам

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Компьютерные технологии обработки больших массивов данных** предполагает использование следующего программного обеспечения и информационных справочных систем:

- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
- доступ в электронную информационно-образовательную среду университета.

Необходимое лицензионное и (или) свободно распространяемое программное обеспечение:

- приложение позволяющее просматривать и воспроизводить медиаконтент PDF-файлов «Adobe Acrobat Reader DC»;
- офисный пакет приложений «LibreOffice», Alt Linux;
- пакета Anaconda для Python версии 3.

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

- система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).
- система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.
- система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения практических занятий - компьютерный класс. Состав оборудования определен в Паспорте компьютерного класса (требуется доступ в Интернет и должен быть установлен пакет Anaconda for Python 3)

Для групповых (индивидуальных) консультаций - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения текущего контроля - аудитория, оснащенная меловой (и) или маркерной доской. Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Компьютерные технологии обработки больших массивов данных**

**Планируемые результаты обучения по дисциплине для формирования компетенции.
Индикаторы и критерии их оценивания**

ОПК.4

Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.4.1 Комбинирует и адаптирует современные информационно-коммуникационные технологии для реализации решения математических задач</p>	<p>Знает и способен использовать технологии обработки текстовых и числовых данных большого объема, может выбрать подходящую структуру данных и запрограммировать алгоритм обработки,</p>	<p align="center">Неудовлетворител</p> <p>Не способен использовать технологии обработки текстовых и числовых данных большого объема, не может выбрать подходящую структуру данных и не способен запрограммировать алгоритм обработки</p> <p align="center">Удовлетворительн</p> <p>Знает и способен использовать технологии обработки текстовых и числовых данных большого объема, может выбрать подходящую структуру данных и запрограммировать алгоритм обработки, действуя строго по образцу.</p> <p align="center">Хорошо</p> <p>Знает и способен использовать технологии обработки текстовых и числовых данных большого объема, может выбрать подходящую структуру данных и запрограммировать алгоритм обработки, допускает 5 и более ошибок, которые может сам исправить после объяснения преподавателя.</p> <p align="center">Отлично</p> <p>Знает и способен использовать технологии обработки текстовых и числовых данных большого объема, может выбрать подходящую структуру данных и запрограммировать алгоритм обработки, допускает не более 5 ошибок, которые сам находит и исправляет.</p>

ОПК.2

Способен совершенствовать и реализовывать новые математические методы решения прикладных задач

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.2.3 Реализует математический метод на языке программирования высокого уровня и/или с помощью специализированных пакетов программ</p>	<p>Умеет использовать библиотеку машинного обучения Sklearn, способен работать с документацией и может применять методы машинного обучения для решения задач анализа данных.</p>	<p>Неудовлетворител Не умеет использовать язык программирования Python и специализированные библиотеки Pandas, Numpy, SciPy, Matplotlib для первичной обработки данных.</p> <p>Удовлетворительн Способен использовать язык программирования Python и специализированные библиотеки Pandas, Numpy, SciPy, Matplotlib для первичной обработки данных, действуя по образцу.</p> <p>Хорошо Умеет использовать язык программирования Python и специализированные библиотеки Pandas, Numpy, SciPy, Matplotlib для первичной обработки данных, но допускает от 5 и более ошибок, которые может исправлять после разъяснений преподавателя.</p> <p>Отлично Умеет использовать язык программирования Python и специализированные библиотеки Pandas, Numpy, SciPy, Matplotlib для первичной обработки данных, но допускает не более 5 ошибок, которые может находить и исправлять самостоятельно.</p>

Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

Максимальное количество баллов : 100

Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 41 до 60

«неудовлетворительно» / «незачтено» менее 41 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
ОПК.2.3 Реализует математический метод на языке программирования высокого уровня и/или с помощью специализированных пакетов программ ОПК.4.1 Комбинирует и адаптирует современные информационно-коммуникационные технологии для реализации решения математических задач	Основы языка программирования Python Письменное контрольное мероприятие	Знание управляющих конструкций языка Python Знание и умение использовать структуры данных языка Python Знание возможностей библиотеки numpy и умение их использовать решения задач линейной алгебрыЗнание возможностей библиотеки scipy и умение их использовать для решения задач оптимизации и математической статистикиЗнание возможностей библиотеки pandas и умение их использовать для работы с базами данныхЗнание возможностей библиотеки matplotlib и умение их использовать для визуализации данных

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<p>ОПК.2.3 Реализует математический метод на языке программирования высокого уровня и/или с помощью специализированных пакетов программ</p> <p>ОПК.4.1 Комбинирует и адаптирует современные информационно-коммуникационные технологии для реализации решения математических задач</p>	<p>Введение в машинное обучение. Обучение с учителем.</p> <p>Защищаемое контрольное мероприятие</p>	<p>Умение использовать линейные модели для решения задачи регрессии с использованием возможностей библиотеки Sklearn. Умение использовать линейные модели для решения задачи классификации с использованием возможностей библиотеки Sklearn. Умение использовать решающие деревья и леса решающих деревьев для решения задачи регрессии с использованием возможностей библиотеки Sklearn.</p> <p>Умение использовать решающие деревья и леса решающих деревьев для решения задачи классификации с использованием возможностей библиотеки Sklearn. Умение использовать методы кросс-валидации для борьбы с переобучением</p> <p>Умение выполнять подбор параметров по сетке</p> <p>Умение работать с документацией библиотеки Sklearn</p>
<p>ОПК.2.3 Реализует математический метод на языке программирования высокого уровня и/или с помощью специализированных пакетов программ</p> <p>ОПК.4.1 Комбинирует и адаптирует современные информационно-коммуникационные технологии для реализации решения математических задач</p>	<p>Практическое применение технологий обработки данных.</p> <p>Итоговое контрольное мероприятие</p>	<p>Умение определять зависимые признаки</p> <p>Умение работать с числовыми признаками</p> <p>Умение работать с бинарными признаками</p> <p>Умение работать с категориальными признаками</p> <p>Умение выбирать подходящие методы и технологии для решения поставленной задачи</p> <p>Умение интерпретировать полученные результаты</p> <p>Умение визуализировать данные и результаты работы</p> <p>Умение документировать работу</p> <p>Умение представить свою работу и отвечать на вопросы во время презентации проекта</p>

Спецификация мероприятий текущего контроля

Основы языка программирования Python

Продолжительность проведения мероприятия промежуточной аттестации: **8 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **12.3**

Показатели оценивания	Баллы
Выполнение контрольной работы	16
Выполнение домашних работ	8
Выполнение практической работы по основам программирования на языке Python	6

Введение в машинное обучение. Обучение с учителем.

Продолжительность проведения мероприятия промежуточной аттестации: **8 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **12.3**

Показатели оценивания	Баллы
Защита реферата по теме "Один из методов машинного обучения: основы алгоритма, область применения, примеры использования с использованием возможностей библиотеки Sklearn.	16
Выполнение домашних заданий	8
Письменные проверочные работы	6

Практическое применение технологий обработки данных.

Продолжительность проведения мероприятия промежуточной аттестации: **8 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **16.4**

Показатели оценивания	Баллы
Защита практической работы анализа данных	32
Выполнение домашних заданий	8