

МИНОБРНАУКИ РОССИИ

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

Авторы-составители: **Шишкин Владимир Андреевич
Радионова Марина Владимировна**

Рабочая программа дисциплины
МАШИННОЕ ОБУЧЕНИЕ ДЛЯ БОЛЬШИХ ДАННЫХ
Код УМК 99864

Утверждено
Протокол №9
от «06» июня 2022 г.

Пермь, 2022

1. Наименование дисциплины

Машинное обучение для больших данных

2. Место дисциплины в структуре образовательной программы

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

Направление подготовки: **01.03.05** Статистика
направленность Анализ больших данных

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Машинное обучение для больших данных** у обучающегося должны быть сформированы следующие компетенции:

01.03.05 Статистика (направленность : Анализ больших данных)

ОПК.4 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Индикаторы

ОПК.4.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности

ПК.3 Способен к информационно-аналитической и организационно-административной поддержке деятельности руководителя организации

Индикаторы

ПК.3.2 Анализирует информацию и готовит информационно-аналитические материалы

4. Объем и содержание дисциплины

Направления подготовки	01.03.05 Статистика (направленность: Анализ больших данных)
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	11
Объем дисциплины (з.е.)	4
Объем дисциплины (ак.час.)	144
Контактная работа с преподавателем (ак.час.), в том числе:	56
Проведение лекционных занятий	28
Проведение практических занятий, семинаров	14
Проведение лабораторных работ, занятий по иностранному языку	14
Самостоятельная работа (ак.час.)	88
Формы текущего контроля	Входное тестирование (1) Защищаемое контрольное мероприятие (2) Итоговое контрольное мероприятие (1)
Формы промежуточной аттестации	Экзамен (11 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Машинное обучение для больших данных

Входной контроль

Проверяется знание основных понятий теории вероятностей и математической статистики, математического анализа, численных методов. Начала программирования.

Введение

Данные и знания. Понятие больших данных. Предварительная подготовка и обработка данных. Программное обеспечение для работы с большими данными.

Тема 1. Регрессионный анализ

Линейные и нелинейные регрессионные модели. Оценка значений параметров регрессионных моделей. Тестирование моделей.

Тема 2. Классификация

Модель бинарного выбора. Линейные и нелинейные разделяющие поверхности. Множественный выбор. Метод Байеса и дерева классификации.

Контрольная точка 1

Контрольная точка по темам "Регрессионный анализ" и "Классификация"

Тема 3. Кластерный анализ

Понятия близости, сходства и эквивалентности. Вычисление расстояния. Иерархические и неиерархические методы кластеризации. Методы k средних, k нечётких средних, Густафсона-Кесселя. Кластеризация, основанная на нечётком отношении альфа квазиэквивалентности. Алгоритмы кластеризации для больших данных.

Тема 4. Факторный анализ

Методы понижения размерности. Метод главных компонент. Сингулярное разложение.

Контрольная точка 2

Контрольная точка по темам "Кластерный анализ" и "Факторный анализ"

Тема 5. Ассоциативные правила

Ассоциативные и секвенциальные правила. Алгоритм Apriori.

Тема 6. Другие методы машинного обучения и анализа данных

Сети векторного квантования (LVQ), Метод опорных векторов (SVM), Бэггинг и случайный лес

Итоговое контрольное мероприятие

Итоговое контрольное мероприятие по всем темам курса

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. Орешков, В. И. Интеллектуальный анализ данных : учебное пособие / В. И. Орешков. — Рязань : Рязанский государственный радиотехнический университет, 2017. — 161 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <https://www.iprbookshop.ru/121842>
3. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/45418>

Дополнительная:

1. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/88752.html>
2. Просто о больших данных:перевод с английского/Д. Гурвиц, А. Ньюджент, Ф. Халпер, М. Кауфман.- Москва:Эксмо,2015, ISBN 978-5-699-85807-1.-400.-Глоссарий: с. 354-368. - Указ.: с. 369-391
3. Адлер, Ю. П. Статистическое управление процессами. «Большие данные» : учебное пособие / Ю. П. Адлер, Е. А. Черных. — Москва : Издательский Дом МИСиС, 2016. — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/64199.html>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

При освоении дисциплины использование ресурсов сети Интернет не предусмотрено.

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

Образовательный процесс по дисциплине **Машинное обучение для больших данных** предполагает использование следующего программного обеспечения и информационных справочных систем:

- презентационные материалы (слайды по темам лекционных и практических занятий);
- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
- доступ в электронную информационно-образовательную среду университета.

Перечень необходимого лицензионного и (или) свободно распространяемого программного обеспечения:

1. свободно распространяемый пакет R
2. офисный пакет прилоений

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

Материально-техническая база обеспечивается наличием:

1. Лекционные занятия - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.
2. Занятий семинарского типа (практические занятия) - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.
3. Лабораторные занятия - Компьютерный класс, оснащенный персональными ЭВМ и соответствующим программным обеспечением. Состав оборудования определен в Паспорте компьютерного класса.
4. Самостоятельная работа - аудитория для самостоятельной работы, оснащенная компьютерной техникой с возможностью подключения к сети «Интернет», обеспеченная доступом в электронную информационно-образовательную среду университета. Помещения Научной библиотеки ПГНИУ.
5. Текущий контроль и промежуточная аттестация - аудитория, оснащенная меловой (и) или маркерной доской.
6. Индивидуальные и групповые консультации - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской или аудитория, оснащенная меловой (и) или маркерной доской.

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
Машинное обучение для больших данных**

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

ОПК.4

Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.4.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p>	<p>Знает критерии выбора информационно-коммуникационных технологий, связанных с машинным обучением на основе обработки больших данных. Умеет использовать выбранные информационно-коммуникационные технологии с учётом требований информационной безопасности. Владеет навыками применения информационно-коммуникационных технологий машинного обучения на основе обработки больших данных.</p>	<p align="center">Неудовлетворител</p> <p>Не знает критерии выбора информационно-коммуникационных технологий, связанных с машинным обучением на основе обработки больших данных. Не умеет использовать выбранные информационно-коммуникационные технологии с учётом требований информационной безопасности. Не владеет навыками применения информационно-коммуникационных технологий машинного обучения на основе обработки больших данных.</p> <p align="center">Удовлетворительн</p> <p>Слабо знает критерии выбора информационно-коммуникационных технологий, связанных с машинным обучением на основе обработки больших данных. Допускает ошибки при использовании выбранных информационно-коммуникационных технологий с учётом требований информационной безопасности. Демонстрирует слабое владение навыками применения информационно-коммуникационных технологий машинного обучения на основе обработки больших данных.</p> <p align="center">Хорошо</p> <p>Знает критерии выбора информационно-коммуникационных технологий, связанных с машинным обучением на основе обработки больших данных. Допускает незначительные ошибки при использовании выбранных информационно-коммуникационных технологий с учётом</p>

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;">Хорошо</p> <p>требований информационной безопасности. Владеет навыками применения информационно-коммуникационных технологий машинного обучения на основе обработки больших данных.</p> <p style="text-align: center;">Отлично</p> <p>Отлично знает критерии выбора информационно-коммуникационных технологий, связанных с машинным обучением на основе обработки больших данных. Умеет использовать выбранные информационно-коммуникационные технологии с учётом требований информационной безопасности. Демонстрирует уверенное владение навыками применения информационно-коммуникационных технологий машинного обучения на основе обработки больших данных.</p>

ПК.3

Способен к информационно-аналитической и организационно-административной поддержке деятельности руководителя организации

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ПК.3.2 Анализирует информацию и готовит информационно-аналитические материалы</p>	<p>Обучающийся способен анализировать информацию и готовить информационно-аналитические материалы. Знает методы машинного обучения, основанные на обработке больших данных. Умеет выполнять предварительную подготовку данных, выбирать подходящие алгоритмы, обрабатывать данные и может анализировать и интерпретировать полученные результаты. Владеет программным обеспечением, предназначенным для работы с большими данными.</p>	<p style="text-align: center;">Неудовлетворител</p> <p>Не знает методы машинного обучения, основанные на обработке больших данных. Не умеет выполнять предварительную подготовку данных, выбирать подходящие алгоритмы, обрабатывать данные и не может анализировать и интерпретировать полученные результаты. Не владеет программным обеспечением, предназначенным для работы с большими данными.</p> <p style="text-align: center;">Удовлетворительн</p> <p>На удовлетворительном уровне знает методы машинного обучения, основанные на обработке больших данных. Допускает ошибки в ходе предварительной подготовки данных, при выборе подходящих алгоритмов, в ходе обработки данных и при</p>

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
		<p style="text-align: center;">Удовлетворительн</p> <p>анализе и интерпретации полученных результатов. Испытывает трудности при работе с программным обеспечением, предназначенным для обработки больших данных.</p> <p style="text-align: center;">Хорошо</p> <p>Знает методы машинного обучения, основанные на обработке больших данных. Допускает незначительные ошибки в ходе предварительной подготовки данных, при выборе подходящих алгоритмов, в ходе обработки данных и при анализе и интерпретации полученных результатов. Испытывает затруднения при работе с программным обеспечением, предназначенным для обработки больших данных.</p> <p style="text-align: center;">Отлично</p> <p>Отлично знает методы машинного обучения, основанные на обработке больших данных. Умеет выполнять предварительную подготовку данных, выбирать подходящие алгоритмы, обрабатывать данные и может анализировать и интерпретировать полученные результаты. В полной мере владеет программным обеспечением, предназначенным для работы с большими данными.</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 43 до 60

«неудовлетворительно» / «незачтено» менее 43 балла

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
Входной контроль	Входной контроль Входное тестирование	Элементы теории вероятностей и математической статистики, математического анализа, численных методов. Начала программирования.
ПК.3.2 Анализирует информацию и готовит информационно-аналитические материалы ОПК.4.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности	Контрольная точка 1 Защищаемое контрольное мероприятие	Основные понятия теории машинного обучения и больших данных. Основные подходы к построению регрессионных моделей данных. Основные методы построения классификаторов для решения задач бинарного и множественного выборов.

Компетенция (индикатор)	Мероприятие текущего контроля	Контролируемые элементы результатов обучения
<p>ПК.3.2 Анализирует информацию и готовит информационно-аналитические материалы</p> <p>ОПК.4.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p>	<p>Контрольная точка 2</p> <p>Защищаемое контрольное мероприятие</p>	<p>Знает понятие обучения без учителя и умеет выбирать подход к построению модели на основе анализа данных. Знает методы кластерного и факторного анализа.</p>
<p>ПК.3.2 Анализирует информацию и готовит информационно-аналитические материалы</p> <p>ОПК.4.2 Ориентируясь на задачи профессиональной деятельности, обоснованно выбирает информационно-коммуникационные технологии и использует их в профессиональной деятельности с учетом требований информационной безопасности</p>	<p>Итоговое контрольное мероприятие</p> <p>Итоговое контрольное мероприятие</p>	<p>Знает методы машинного обучения на основе обработки больших данных. Знает методы обучения с учителем и без учителя. Знает методы построения ассоциативных и секвенциальных правил. Знает методы обработки нечисловой информации (текст, изображение, видео).</p>

Спецификация мероприятий текущего контроля

Входной контроль

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **0**

Проходной балл: **0**

Показатели оценивания	Баллы
Знает и умеет использовать основные понятия программирования. Знает основные структуры данных и программные конструкции.	4
Знает и умеет использовать основные понятия и методы теории вероятностей и математической статистики.	3
Знает и умеет использовать основные понятия и методы математического анализа. Знает основы численных методов.	3

Контрольная точка 1

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

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

Проходной балл: **13**

Показатели оценивания	Баллы
Знает подходы к построению регрессионных моделей данных.	11
Знает методы построения классификаторов при решении задач бинарного и множественного выборов.	11
Знает понятия машинного обучения и больших данных. Знает методы подготовки больших данных к дальнейшему анализу.	8

Контрольная точка 2

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

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

Проходной балл: **13**

Показатели оценивания	Баллы
Знает методы кластерного анализа. Умеет использовать алгоритмы иерархической и неиерархической кластеризации. Знает методы нечёткой кластеризации.	16
Знает методы факторного анализа и умеет использовать их для сжатия данных.	10
Знает понятие обучения без учителя. Умеет выбирать подход к построению модели на основе анализа данных.	4

Итоговое контрольное мероприятие

Продолжительность проведения мероприятия промежуточной аттестации: **1 часа**

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

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

Проходной балл: **17**

Показатели оценивания	Баллы
Знает методы построения часто встречающихся наборов для пространственных и временных данных, алгоритм Apriori и его модификации. Знает методы построения и оценки качества ассоциативных и секвенциальных правил.	20
Знает методы анализа и моделирования нечисловой информации (текст, изображение, видео).	10
Знает понятия обучения с учителем и без учителя. Умеет проводить предварительную обработку данных, а также выбирать тип модели на основе анализа данных.	5
Знает методы построения моделей для маркированных и немаркированных данных: регрессионный анализ, построение классификаторов, кластерный анализ, факторный анализ.	5