Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования «Северо-Осетинский государственный университет

имени Коста Левановича Хетагурова»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«**Машинное обучение**»

Направление подготовки:

**01.03.02 Прикладная математика и информатика**

Профиль: **«Программирование, анализ данных**

**и математическое моделирование»**

Квалификация (степень) выпускника – бакалавр

Форма обучения очная

Год начала подготовки 2024

Владикавказ

2024

Рабочая программа

*обсуждена и утверждена* на заседании кафедры прикладной математики и информатики (протокол № 6 от 27.02.2024 г.);

*одобрена* советом факультета математики и компьютерных наук (протокол № 6 от 01.03.2024 г.);

*утверждена* в составе Основной профессиональной образовательной программы понаправлению подготовки 01.03.02 «Прикладная математика и информатика», профиль: «Программирование, анализ данных и математическое моделирование», год начала подготовки 2024 (решение ученого совета от 28.03.2024, протокол № 8).

Составители: доцент кафедры прикладной математики и информатики, к.ф.-м.н. Гутнова А.К., доцент кафедры прикладной математики и информатики, к.т.н. Калиниченко А.В.

**1. Структура и общая трудоемкость дисциплины**

Общая трудоемкость дисциплины составляет 5 з.е. (180 ч.).

|  |  |
| --- | --- |
|  | форма обучения |
| Курс | 4 |
| Семестр | 7 |
| Лекции | - |
| Практические занятия | - |
| Лабораторные занятия | 34 |
| Консультации |  |
| Итого аудиторных занятий | 34 |
| Самостоятельная работа | 110 |
| Курсовая работа | - |
| Зачет | - |
| Экзамен | 36 |
| Общее количество часов | 180 |

**2. Цели изучения дисциплины**

Цель изучения дисциплины – ознакомление студентов с теоретическими основами и основными принципами машинного обучения — а именно, с классами моделей (линейные, логические, нейросетевые), метриками качествами и подходами к подготовке данных, формирование у студентов практических навыков работы с данными и решения прикладных задач анализа данных.

**3. Место дисциплины в структуре ОПОП:**

Б1.В.ДВ.02.01. Блок 1. Дисциплины (модули). Часть, формируемая участниками образовательных отношений.

Для освоения данной дисциплины полезны знания, умения, навыки и компетенции, формируемые предшествующими дисциплинами «Линейная алгебра» «Аналитическая геометрия», «Математический анализ», «Теория вероятностей и математическая статистика», «Алгоритмы и структуры данных», «Основы и методология программирования».

**4. Требования к результатам освоения дисциплины**

В результате освоения дисциплины обучающийся должен обладать следующими компетенциями (результатами освоения образовательной программы):

Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);

Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде (УК-3);

Способен проводить научно-исследовательские разработки по отдельным разделам темы в области прикладной математики и информационных технологий (ПК-1);

Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники (ПК-2).

Взаимосвязь планируемых результатов обучения по дисциплине с формируемыми компетенциями ОПОП:

|  |  |  |  |
| --- | --- | --- | --- |
| **Компетенции** | **Планируемые результаты обучения, соответствующие формируемым компетенциям ОПОП** | | |
| **Код и формулировка** | **Знать:** | **Уметь** | **Владеть:** |
| УК-1  Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач | – способы поиска, анализа и синтеза информации для решения задач анализа данных методами машинного обучения. | – осуществлять синтез анализ и информации при решении задач методами машинного обучения;  – умеет выбирать методы машинного обучения, подходящие для решения конкретной задачи и выбирать оптимальный из них | – имеет практический опыт решения задач машинного обучения, с применением системного подхода. |
| УК-3  Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде | различные приемы и способы социализации личности и социального взаимодействия при работе над командным проектом | строить отношения с членами команды в процессе работы над проектом | практический опыт участия в командной работе, в социальных проектах, распределения ролей в условиях командного взаимодействия |
| ПК-1  Способен проводить научно-исследовательские разработки по отдельным разделам темы в области прикладной математики и информационных технологий | – методы сбора, обработки и анализа научно-технической информации и представления результатов исследований. | – осуществлять поиск, отбор и изучение научной литературы в области анализа данных и машинного обучения. | – практический опыт научно-исследовательской деятельности в области анализа данных и машинного обучения. |
| ПК-2  Способен создавать и исследовать новые математические модели в естественных науках, промышленности и бизнесе, с учетом возможностей современных информационных технологий и программирования и компьютерной техники | – методы разработки моделей машинного обучения;  – метрики качества моделей машинного обучения. | – использовать современные информационные технологии, методы программирования и компьютерную технику для создания и исследования моделей машинного обучения и анализа данных. | – практический опыт разработки и исследования моделей машинного обучения. |

**5. Содержание и учебно-методическая карта дисциплины**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **№ темы** | **Наименование тем (вопросов), изучаемых по данной дисциплине** | **Занятия** | | **Самостоятельная работа студентов** | | **Формы контроля** | **Литература** |
| **л** | **пр.** | **Содержание** | **Часы** |
|  | Биологический нейрон, модель МакКаллока-Питтса как линейный классификатор. Функции активации. |  | 2 | Нейронные сети в обработке изображений. Фильтры. Сверточные слои. | 14 | Опрос, доклад, лабораторная работа | [1-7] |
|  | Проблема полноты. Задача исключающего или. Полнота двухслойных сетей в пространстве булевых функций.  Алгоритм обратного распространения ошибок. |  | 4 | Нейронные сети и обучение представлений. Обработка последовательностей. | 16 | Опрос, доклад, лабораторная работа | [1-7] |
|  | Быстрые методы стохастического градиента: Поляка, Нестерова, AdaGrad, RMSProp |  | 4 | Методы AdaDelta, Adam, Nadam. | 20 | Опрос, доклад, лабораторная работа | [1-7] |
|  | Метод случайных отключений нейронов (Dropout). Интерпретации Dropout. Обратный Dropout и L2-регуляризация. |  | 4 | Проблемы: затухающие и взрывающиеся градиенты, невыпуклость функции потерь. | 20 | Опрос, доклад, лабораторная работа | [1-7] |
|  | Функции активации ReLU и PReLU. |  | 4 |  |  | Опрос, доклад, лабораторная работа | [1-7] |
|  | Эвристики для формирования начального приближения. Метод послойной настройки сети. |  | 4 |  |  | Опрос, доклад, лабораторная работа | [1-7] |
|  | Подбор структуры сети: методы постепенного усложнения сети, оптимальное прореживание нейронных сетей (optimal brain damage). |  | 4 |  |  | Опрос, доклад, лабораторная работа | [1-7] |
|  | Глубокие нейронные сети  а) Свёрточные нейронные сети (CNN) для изображений. Свёрточный нейрон. Pooling нейрон. Выборка размеченных изображений ImageNet.  б) ResNet: остаточная нейронная сеть (residual NN). Сквозные связи между слоями (skip connection).  в) Рекуррентные нейронные сети (RNN). Обучение рекуррентных сетей: Backpropagation Through Time (BPTT).  г) Сети долгой кратковременной памяти (Long short-term memory, LSTM). |  | 4 | Рекуррентные сети Gated Recurrent Unit (GRU) и Simple Recurrent Unit (SRU). | 20 | Опрос, доклад, лабораторная работа | [1-7] |
|  | Нейронные сети с обучением без учителя.  а) Нейронная сеть Кохонена. Конкурентное обучение, стратегии WTA и WTM.  б) Автокодировщик. Линейный AE, SAE, DAE, CAE, RAE, VAE, AE для классификации, многослойный AE.  в) Пред-обучение нейронных сетей (pre-training).  г) Перенос обучения (transfer learning). |  | 4 | Многозадачное обучение (multi-task learning).  Самостоятельное обучение (self-supervised learning). | 20 | Опрос, доклад, лабораторная работа | [1-7] |
|  | **Итого** | - | **34** |  | **110** |  |  |

**6. Образовательные технологии**

Согласно учебному плану при преподавании дисциплины используются традиционные образовательные технологии: лабораторные работы и самостоятельная работа студентов. Также при проведении занятий и самостоятельной работе студентов могут быть использованы:

**–** *интерактивные технологии* («мозговой штурм», дебаты, презентационный метод, работа в парах, работа в группах, деловая игра);

**–** *технологии контекстного обучения* **–** система дидактических форм, методов и средств, направленная на моделирование содержания будущей профессиональной деятельности специалиста (анализ конкретных ситуаций, методы работы с информационными базами данных, деловая игра и др.);

– *технологии электронного обучения* (реализуется при помощи электронной образовательной среды СОГУ при использовании ресурсов ЭБС, при проведении автоматизированного тестирования и т. д.).

Все виды учебной работы могут проводиться дистанционно на основании локальных нормативных актов СОГУ.

В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по индивидуальной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины может осуществляться через индивидуальные консультации преподавателя очно, в часы консультаций, по электронной почте или с использованием ЭИОС СОГУ.

**7. Учебно-методическое обеспечение самостоятельной работы**

Самостоятельная работа проводится с целью:

– систематизации и закрепления полученных теоретических знаний и практических умений обучающихся студентов;

– углубления и расширения теоретических знаний;

– формирования умений использовать нормативную, правовую, справочную документацию и специальную литературу;

– формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;

– развития исследовательских навыков и умений.

По дисциплине предусмотрены следующие виды самостоятельной работы студентов:

– самостоятельное повторение и изучение теоретического материала;

– выполнение домашних заданий;

– подготовка доклада/конспекта по теме, вынесенной на самостоятельное изучение;

– подготовка к выполнению лабораторных работ;

– подготовка к промежуточной аттестации (экзамену).

Содержание, трудоемкость и формы контроля внеаудиторной самостоятельной работы содержатся в разделе 5.

**8. Оценочные средства для текущего контроля успеваемости,**

**рубежной и промежуточной аттестации по итогам освоения дисциплины**

Основными *формами текущего контроля* по дисциплине являются:

– устный опрос на практическом занятии;

– проект по теме, вынесенной на самостоятельное изучение;

– проверка домашних заданий, лабораторных работ.

Форма *рубежного контроля:* контрольная работа.

Формы *промежуточной аттестации:* экзамен.

**8.1. Формы контроля и критерии оценивания**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Этап** | **Форма**  **контроля** | **Критерии оценивания (процент от максимального кол-ва баллов)** | | | |
| **86-100 %** | **71–85%** | **50–70%** | **Менее 50%** |
| **отлично / зачет** | **хорошо / зачет** | **удовлетворительно / зачет** | **неудовлетворительно / незачет** |
| *1. Текущий контроль (max 20 баллов за один модуль)* | | | | | |
|  |  | 15–17 баллов | 12–14 баллов | 8–11 баллов | 0–7 баллов |
|  | Текущая работа в течение модуля  (мах 17б.) | Студент активно работает на занятиях, превосходно выполняет все задания преподавателя. | Студент активно работает на занятиях, хорошо выполняет задания преподавателя. | Студент недостаточно активно работает на занятиях, удовлетворительно выполняет задания преподавателя. | Студент недостаточно активно работает на занятиях, неудовлетворительно выполняет задания преподавателя. |
|  |  | 3 балла | 2 балла | 1 балл | 0 баллов |
|  | Реферат /  доклад / конспект  (мах 3б.) | Тема полностью раскрыта. Превосходное владение материалом. Высокий уровень самостоятельности, логичности, аргументированности. Превосходный стиль изложения. | Тема в основном раскрыта. Хорошее владение материалом. Средний уровень самостоятельности, логичности, аргументированности. Хороший стиль изложения. | Тема частично раскрыта. Удовлетворительное владение материалом. Низкий уровень самостоятельности, логичности, аргументированности. Удовлетворительный стиль изложения. | Тема не раскрыта. Неудовлетворительное владение материалом. Недостаточный уровень самостоятельности, логичности, аргументированности. Неудовлетворительный стиль изложения. |
| *2. Рубежный контроль (15 б. за один модуль)* | | | | | |
|  | тест /  контрольная работа | Количество баллов за выполнение каждого задания указываются в тесте / контрольной работе. | | | |
| *3. Промежуточная аттестация по дисциплине (max число баллов – в соответствии с действующим локальным нормативным актом)* | | | | | |
|  |  | Критерии оценивания (процент от максимального кол-ва баллов) | | | |
|  |  | 86–100 % | 71–85 % | 50–70 % | 0–49 % |
|  | Зачет /  Экзамен | Дан полный, развернутый ответ на поставленный вопрос. Ответ формулируется в терминах науки, изложен литературным языком, логичен, доказателен, демонстрирует авторскую позицию студента. | Дан полный ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные связи. Но допущены незначительные ошибки, исправленные студентом с помощью «наводящих» вопросов преподавателя. | Дан недостаточно полный ответ. Студент не способен самостоятельно выделить существенные и несущественные признаки и причинно-следственные связи. Речевое оформление требует поправок, коррекции. | Не получены ответы по базовым вопросам дисциплины или дан неполный ответ и допущены грубые ошибки. Речь неграмотная. Уточняющие вопросы преподавателя не приводят к коррекции ответа студента не только на поставленный вопрос, но и на другие вопросы дисциплины. |

Пересчет полученной суммы баллов по дисциплине в оценку производится в соответствии с действующим локальным нормативным актом.

**8.2. Примерный вариант 1 рубежной контрольной работы (15 баллов)**

Построить случайный лес, используя RandomForestClassifier. Вывести значимые признаки в порядке убывания значимости. Датасет доступен по адресу: <https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset>

**8.3. Примерный вариант 2 рубежной контрольной работы (теста) (15 баллов)**

Принять участие в соревновании от MLCourse.AI на ODS: <https://ods.ai/tracks/linear-models-spring22/competitions/alice> и попасть в 30% топ-участников. (15б)

**8.4. Вопросы к экзамену**

1. Основные понятия и определения машинного обучения
2. Классификация алгоритмов машинного обучения.
3. Этапы машинного обучения.
4. Задачи обучения с учителем.
5. Постановка задач классификации, регрессии, прогнозирования и ранжирования.
6. Современные методы решения задач распознавания образов.
7. Рекомендательные системы.
8. Метод опорных векторов.
9. Метод наименьших квадратов.
10. Задачи обучения без учителя.
11. Постановка задач кластеризации, поиска ассоциативных правил, фильтрации выбросов, построение доверительной области, сокращение размерности, заполнение пропущенных значений.
12. Этапы проведения кластерного анализа.
13. Методы кластерного анализа.
14. Алгоритмы генерации ассоциативных правил. Фильтрация данных.
15. Современные нейросетевые технологии.
16. Классификация архитектур нейронных сетей.
17. Нейронные сети глубокого обучения.
18. Рекуррентные нейронные сети.
19. Сверточные нейронные сети.
20. Нейронные сети долгой краткосрочной памяти.
21. Задачи обучения с подкреплением.
22. Основные понятия и этапы работы генетических алгоритмов.
23. Основные понятия агентного подхода. Многоагентные системы.
24. Алгоритмы SARSA и Q-learning.
25. Обработка естественного языка. Особенности обработки естественного языка.
26. Библиотеки и фреймворки для распознавания речи, анализа и генерирования текста, синтеза речи.

**8.5. Примерные темы лабораторных работ**

1. Задачи обучения с учителем.
2. Современные методы решения задач распознавания образов.
3. Рекомендательные системы.
4. Разработка когнитивной системы распознавания образов.
5. Обработка естественного языка.
6. Библиотеки и фреймворки для распознавания речи, анализа и генерирования текста, синтеза речи.
7. Создание интеллектуального чат-бота.
8. Задачи обучения без учителя.
9. Методы кластерного анализа.
10. Определение кластеров в наборах данных.
11. Современные нейросетевые технологии.
12. Рекуррентные нейронные сети.
13. Сверточные нейронные сети.
14. Восстановление искаженных образов с использованием рекуррентной нейронной сети.
15. Задачи обучения с подкреплением.
16. Основные понятия и этапы работы генетических алгоритмов.
17. Решение задачи поиска экстремума функции с помощью генетического алгоритма.
18. Задачи обучения с подкреплением.
19. Алгоритмы SARSA и Q-learning.
20. Конструирование функции агента.

**9. Учебно-методическое и информационное обеспечение дисциплины**

**а) основная литература:**

1. Л. Рамальо. Python. К вершинам мастерства: ДМК Пресс, 2016//ЭБС издательства «Лань».- URL: <https://e.lanbook.com/book/93273>.
2. Е.А. Черткова Статистика. Автоматизация обработки информации: Учебное пособие Юрайт, 2018//ЭБС Юрайт. – URL: <https://www.biblio-online.ru/book/0CBA0F5B-1227-46F3-8C8E-D9BAB4AC306A>.
3. Воронцов К.В. Машинное обучение Школа Анализа данных Яндекс. МФТИ. национальный открытый университет Интуит, 2015 – Режим доступа: свободный – URL: <https://www.intuit.ru/studies/courses/13844/1241/info>

**б) дополнительная литература:**

1. Хахаев И. А. Практикум по алгоритмизации и программированию на Python: курс. М.: Национальный Открытый Университет «ИНТУИТ», 2016//ЭБС «Университетская библиотека online». – URL: <http://biblioclub.ru/index.php?page=book&id=429256>.
2. Бессмертный И.А., Нугуманова А.Б., Платонов А.В. Интеллектуальные системы: Учебник и практикум для академического бакалавриата Юрайт, 2018 // ЭБС "Юрайт". – URL: <https://urait.ru/book/intellektualnye-sistemy-469867>.
3. Крутиков В. Н., Мешечкин В. В. Анализ данных: учебное пособие Кемеровский государственный университет, 2014 // ЭБС "Университетская библиотека online".
4. Яковлев, В. Б.  Статистика. Расчеты в Microsoft Excel: учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2017. — 353 с. — (Университеты России). — ISBN 978-5-534-01672-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/400278>.

***в) Профессиональные базы данных и другие интернет-ресурсы:***

– Kaggle: система организации конкурсов по анализу данных, а также социальная сеть специалистов по обработке данных и машинному обучению (<https://www.kaggle.com/>);

– Яндекс практикум: курсы программирования (платные и бесплатные) (Основы программирования, Основы Python-разработки, Основы Go, Python-разработчик, Разработчик С++, Android-разработчик, Мидл Python-разработчик, Асинхронное программирование на Python и др.) (<https://practicum.yandex.ru/catalog/programming/>);

– Stepik: курсы по программированию (платные и бесплатные) (<https://stepik.org/catalog>);

– ЭБС «Научная электронная библиотека eLibrary.ru» (<http://www.elibrary.ru>);

– ЭБС «Университетская библиотека onLine» (<http://www.biblioclub.ru>);

– ЭБС «Юрайт» (<http://www.urait.ru/>);

– Универсальная база данных «ИВИС» ([htpps:/eivis.ru/](about:blank));

– ИС «Национальнаяэлектронная библиотека (НЭБ)»(<https://rusneb.ru/>).

**10. Материально-техническое обеспечение дисциплины**

Для проведения лабораторных занятий используются:

– учебные аудитории для проведения лекционных занятий, занятий семинарского типа, текущего контроля, промежуточной аттестации, оборудованные аудиторной мебелью, доской (меловой, маркерной или интерактивной), компьютером или ноутбуком с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду СОГУ, мультимедийным проектором, экраном;

– компьютерный класс (корпус 10, ауд. №505, 506, 600, 601, 605, 606), оборудованный аудиторной мебелью, доской (меловой, маркерной или интерактивной), компьютерами или ноутбуками с возможностью подключения к сети «Интернет» и доступом к электронной информационно-образовательной среде СОГУ, мультимедийным проектором, экраном.

*Лицензионное программное обеспечение:*

1. Windows 10 Pro for Workstations, (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
2. Office Standard 2016 (№ 4100072800 Microsoft Products (MPSA) от 04.2016г);
3. Система поиска текстовых заимствований «Антиплагиат ВУЗ».

*Перечень ПО в свободном доступе:* Kaspersky Free; WinRar; Google Chrome; Yandex Browser; Opera Browser; Acrobat Reader; MOODLE.

Помещение для самостоятельной работы студентов: Зал электронных ресурсов Научной библиотеки СОГУ (корпус 6, кабинет № 1.8), укомплектован специализированной мебелью (рабочие места студентов), необходимыми техническими средствами обучения: компьютеры, принтер, возможность подключения к сети «Интернет», доступ в электронную информационно-образовательную среду СОГУ.