Федеральное государственное бюджетное образовательное учреждение высшего образования

«ОМСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

«Утверждаю»

Проректор по ОД

А.С. Полынский

« » 2021 год

Р А Б О Ч А Я П Р О Г Р А М М А

по дисциплине

**«Машинное обучение»**

|  |  |
| --- | --- |
| **02.03.02** | **«*Фундаментальная информатика и информационные технологии*»**  **(«*Технологии искусственного интеллекта*»)** |
|  |  |
| **02.03.03** | **«*Математическое обеспечение и администрирование информационных систем*»**  **(«*Технологии больших данных*»)** |

Разработана в соответствии с ФГОС ВО 3++, ООП по направлениям подготовки бакалавриата 02.03.02 «Фундаментальная информатика и информационные технологии» направленности «Технологии искусственного интеллекта», 02.03.03 «Математическое обеспечение и администрирование информационных систем» направленности «Технологии больших данных».

|  |  |  |
| --- | --- | --- |
| Программу составил: | | |
| канд. физ.-мат. наук, доц., доц. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | / Канева О.Н./ |
| ученая степень, звание, должность |  | ФИО |
|  | «\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_ | 2021 г |

|  |  |  |
| --- | --- | --- |
| Обсуждена на заседании кафедры «Прикладная математика и фундаментальная информатика», | | |
| протокол от «\_\_\_\_\_\_» \_\_\_\_\_\_ 2021 г. №\_\_\_\_. |  |  |
| Зав. кафедрой «Прикладная математика и фундаментальная информатика» | | |
|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | / Зыкина А.В. / |
|  |  | ФИО |
|  | «\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_ | 2021 г. |

|  |  |  |
| --- | --- | --- |
| Руководитель ООП по направлению подготовки бакалавриата 02.03.02 «Фундаментальная информатика и информационные технологии» направленности «Технологии искусственного интеллекта». | | |
| проф., д-р физ.-мат. наук, проф. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | /Зыкина А.В./ |
| ученая степень, звание, должность |  | ФИО |
|  | «\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_2021 г. |  |

|  |  |  |
| --- | --- | --- |
| Руководитель ООП по направлению подготовки бакалавриата 02.03.03 «Математическое обеспечение и администрирование информационных систем» направленности «Технологии больших данных». | | |
| проф., д-р физ.-мат. наук, проф. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | /Зыкина А.В./ |
| ученая степень, звание, должность |  | ФИО |
|  | «\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_2021 г. |  |

Рабочая программа разработана для

|  |  |  |  |
| --- | --- | --- | --- |
| Год набора | Начальник УМУ (подпись, дата) | Год набора | Начальник УМУ (подпись, дата) |
| 2021 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## Цели освоения дисциплины

Целью дисциплины «Машинное обучение» является изучение алгоритмов, которые самостоятельно обучаются и прогнозируют ситуацию на основе имеющихся данных.

Основные задачи дисциплины:

1. изучить типы машинного обучения;
2. изучить основные алгоритмы машинного обучения;
3. научиться строить конвейеры из уже реализованных в библиотеках алгоритмов и реализовывать собственные алгоритмы для обработки и анализа данных средствами Python.

## Место дисциплины в структуре ООП

Дисциплина «Машинное обучение» изучается в 4 семестре и относится к обязательной части учебного плана ООП по направлениям подготовки бакалавров 02.03.02 «Фундаментальная информатика и информационные технологии», 02.03.03 «Математическое обеспечение и администрирование информационных систем».

Для успешного изучения дисциплины необходимы знания и умения, полученные при изучении предшествующих дисциплин:

1. Математический анализ;
2. Дополнительные главы математического анализа;
3. Оптимизация и принятие решений;
4. Алгебра.

Дисциплины, изучаемые одновременно:

1. Теория вероятностей и математическая статистика.
2. Проектная деятельность.

Освоение данной дисциплины является необходимой основой для изучения последующих дисциплин по направлению подготовки бакалавров 02.03.02 «Фундаментальная информатика и информационные технологии»:

1. Обработка естественного языка;
2. Компьютерное зрение;
3. Анализ временных рядов;
4. Моделирование систем и процессов;
5. Математические методы прогнозирования;
6. Визуализация многомерных данных;
7. Разработка интеллектуальных систем;

и по направлению подготовки бакалавров 02.03.03 «Математическое обеспечение и администрирование информационных систем»:

1. Прикладные задачи анализа данных;
2. Продвинутые методы работы с данными;
3. BI решения и многомерная модель данных;
4. Большие данные.

## Требования к результатам освоения дисциплины

В результате освоения дисциплины «Машинное обучение» должны быть сформированы компетенции, приведенные в таблице 3.1. Обучающийся должен демонстрировать освоение указанными компетенциями по дескрипторам «знания, умения, владения», соответствующим тематическим модулям дисциплины и применимым в их последующем обучении и профессиональной деятельности.

Таблица 3.1

|  |  |  |
| --- | --- | --- |
| **Код компетенции** | **Код индикатора** | **Проектируемые результаты освоения дисциплины** |
| **ПК-3.** Способен разрабатывать и реализовывать базы данных и хранилища данных для систем поддержки принятия решений. | **ИД-2ПК-3.** Проводит обследования, консолидацию, обработку и анализ информации из различных источников. | **Знать:** методологии и технологии Data Science.  **Уметь:** разрабатывать и реализовывать проект Data Science.  **Владеть:** навыками использования необходимого для реализации проекта Data Science инструментария и инфраструктур. |
| **ИД-3ПК-3.**Извлекает имеющую смысл информацию из данных с помощью методов машинного обучения. | **Знать:** различные методы и алгоритмы машинного обучения; критерии качества моделей машинного обучения; последовательность разработки моделей машинного обучения.  **Уметь:** применять алгоритмы и методы машинного обучения.  **Владеть:** навыками использования библиотек Python для быстрого сбора конвейера для решения задач анализа данных. |

## Объем дисциплины и виды учебной работы в часах и зачетных единицах

Таблица 4.1

## Очная форма обучения

|  |  |  |
| --- | --- | --- |
| Вид занятий | Всего  (час) | С е м е с т р ы |
| **4** |
| **Всего аудиторных занятий:** | **72** | **72** |
| Практические занятия | 36 | 36 |
| Лабораторные работы | 36 | 36 |
| **Самостоятельная работа:** | **72** | **72** |
| СРС | 52 | 52 |
| КСР (РГР) | 20 | 20 |
| **Количество часов на экзамен** | **36** | **36** |
| **Всего по дисциплине** | **180/5** | **180** |
| Вид аттестации за семестр (зачет,  дифференцированный зачет, экзамен) |  | экзамен |

## Содержание дисциплины по разделам и видам учебных занятий

* 1. **Содержание дисциплины по разделам**
     + 1. Процесс Data Science.
       2. Предобработка данных.
       3. Обучение с учителем.
       4. Обучение без учителя.
       5. Обучение с подкреплением.
       6. Разработка прикладных решений.
  2. **Содержание практических занятий**

Содержание практических занятий приведено в таблице 5.1. Номера разделов практических занятий соответствуют номерам разделов содержания дисциплины.

Таблица 5.1

| **№** | **Содержание практических занятий** | Форма  обучения |
| --- | --- | --- |
| Очная |
| Кол-во часов |
| **1** | **Раздел 1. Процесс Data Science** | **2** |
| Процесс Data Science. Этапы, инструменты и инфраструктуры для реализации процесса Data Science. Место машинного обучения в процессе Data Science. Классификация задач ML. Этапы построения систем машинного обучения |
| **2** | **Раздел 2. Предобработка данных** | **4** |
| Предобработка данных: решение проблемы пропущенных данных, обработка категориальных данных, приведение признаков к одинаковой шкале, отбор содержательных признаков: одномерный отбор признаков, жадные методы отбора признаков, отбор признаков на основе моделей. |
| **3** | **Раздел 3. Обучение с учителем** | **18** |
| Регрессионный анализ: постановка задачи, метрики качества, линейная модель, методы оценки параметров линейной регрессии, регуляризация, спрямляющие пространства. |
| Задача классификации: постановка задачи, метрики качества, линейная модель, методы опорных векторов. Деревья принятия решений. Нейронные сети. |
| Методы оценки моделей: отложенная выборка и кросс-валидация, отладка алгоритмов при помощи кривой обучения и проверочной кривой Настройка машиннообучаемых моделей методом сеточного поиска, обзор других метрик оценки качества. |
| Композиции простых алгоритмов. Случайный лес. Бустинг: градиентный бустинг, стохастический градиентный бустинг. |
| **4** | **Раздел 4. Обучение без учителя** | **12** |
| Кластерный анализ. Методы кластеризации: метод K средних (K-Means), Expectation Maximization (EM-алгоритм), агломеративная иерархическая кластеризация, графовые методы кластеризации, методы, основанные на плотности. Оценка качества и рекомендации по решению задачи кластеризации. |
| Понижение размерности: метод главных компонент. |
| [Задача обнаружения аномалий](https://www.coursera.org/lecture/unsupervised-learning/zadacha-obnaruzhieniia-anomalii-N3yt5). Параметрическое восстановление плотности. Непараметрическое восстановление плотности. Задача визуализации. Многомерное шкалирование. Метод t-SNE. |
| Постановка задачи тематического моделирования. Базовые тематические модели и EM-алгоритм. Регуляризация тематических моделей. Мультимодальные тематические модели. Внутренние критерии качества тематических моделей. Внешние критерии качества тематических моделей. |
|  | **ИТОГО ЧАСОВ** | **36** |

* 1. **Содержание лабораторных занятий**

Цель лабораторного практикума – закрепление теоретического материала дисциплины, овладение методами решения задач. Содержание лабораторных работ приведено ниже в таблице 5.2. Номера разделов лабораторных работ соответствуют номерам разделов содержания дисциплины.

Таблица 5.2

| **№** | **Содержание лабораторного практикума** | Форма обучения |
| --- | --- | --- |
| Очная |
| Кол-во часов |
| **1** | **Раздел 2. Предобработка данных** |  |
| Лабораторная работа №1. Подготовка рабочего окружения | **2** |
| Лабораторная работа №2. Разведочный анализ данных | **2** |
| **2** | **Раздел 3. Обучение с учителем** |  |
| Лабораторная работа №3. Классификация | **2** |
| Лабораторная работа №4. Регрессия | **2** |
| Лабораторная работа №5. Композиция алгоритмов | **2** |
| Лабораторные работы №6-8. Нейронные сети. Глубокое обучение. Архитектуры. | **6** |
| Лабораторные работы №. 9-11 Обработка изображений с помощью методов глубокого обучения. | **6** |
| **3** | **Раздел 5. Обучение с подкреплением** |  |
| Лабораторные работы №12-14. Обучение с подкреплением | **6** |
| **4** | **Раздел 6. Разработка прикладных решений** |  |
| Лабораторные работы №15-17. Разработка прикладных решений. | **6** |
| Лабораторная работа №18. Экспертиза разработанных прикладных решений. | **2** |
|  | **ИТОГО ЧАСОВ** | **36** |

## Самостоятельная работа студентов

Самостоятельная работа направлена на закрепление и углубление полученных теоретических и практических знаний, развитие навыков практической работы.

Объем СРС и распределение по видам учебных работ в часах представлены в таблице 5.3.

Таблица 5.3

| **Вид СРС** | Количество часов |
| --- | --- |
| Семестр |
| **4** |
| **СРС** | **52** |
| Работа с теоретическим материалом, самостоятельное изучение разделов дисциплины; поиск и обзор литературы и электронных источников; чтение и изучение учебников и учебных пособий. | 22 |
| Подготовка к лабораторным работам. | 30 |
| **КСР** | **20** |
| Выполнение РГР | 20 |
| **ИТОГО по дисциплине** | **72** |

1. **Образовательные технологии**

При реализации дисциплины используются различные образовательные технологии:

– традиционные образовательные технологии: практические занятия, лабораторные работы, самостоятельная работа и индивидуальные задания;

– инновационные образовательные технологии: элементы дистанционного обучения.

Самостоятельная работа студентов предусматривает выполнение расчетно-графической работы. Кафедра имеет право изменять последовательность прохождения отдельных тем. Преподаватель может выбирать технологию и метод обучения.

1. **Методические указания по освоению дисциплины**

Материалы, используемые для проведения занятий, соответствуют учебному плану, рабочей программе дисциплины и позволяют полностью освоить заданные компетенции. Контактная работа проводится в виде расчетно-графической работы – разработка и реализация проекта индивидуально или в команде. Вид и форма материала для практических занятий определяется преподавателем. Материал для практических занятий содержит: конспект с необходимым для выполнения практических заданий содержанием, задачи, задания и варианты их решения.

Перечень методических материалов по освоению дисциплины доводится до сведения обучающихся на первом занятии по дисциплине.

## Оценочные средства

## Оценочные средства позволяют оценить знания, умения, владение и уровень приобретенных компетенций. Оценка качества освоения программы дисциплины «Машинное обучение» включает текущий контроль успеваемости, промежуточную аттестацию.

## Текущий контроль успеваемости обучающихся предназначен для проверки хода и качества усвоения учебного материала, качества и своевременности выполнения обязательных заданий (учебных работ), включенных в рабочую программу дисциплины, оценки степени (уровня) достижения учебных целей по формированию компетенций, отнесенными к изучаемым темам, разделам по дисциплине.

## Для фиксации результатов текущего контроля успеваемости обучающихся в семестре, применяется балльная система. Шкала текущих балльных оценок за выполнение и защиту обязательных заданий, и степень (уровня) достижения учебных целей по формированию компетенций по всем формам текущего контроля разрабатывается ведущим преподавателем кафедры, за которой закреплена дисциплина. Критерии оценки по каждому виду обязательного задания доводятся до сведения обучающихся на первом занятии по дисциплине.

## Промежуточная аттестация обучающихся фиксирует окончательные результаты обучения по дисциплине.

## Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

К промежуточной аттестации студентов по дисциплине «Машинное обучение» могут привлекаться в качестве внешних экспертов: представители работодателей, преподаватели последующих дисциплин.

Оценочные средства по дисциплине «Машинное обучение» включают:

* электронное портфолио обучающегося;
* контрольные вопросы по дисциплине;
* комплекты заданий для выполнения расчетно-графической работы;
* экзаменационные билеты.
  1. **Расчетно-графическая работа**

По дисциплине предусмотрена расчетно-графическая работа в рамках которой студенты разрабатывают алгоритмы на основе методов машинного обучения для решения задач анализа данных из различных прикладных областей. Приветствуется вариант формулировки темы РГР по согласованию со студентами в рамках задач выполняемого проекта по дисциплине «Проектная деятельность».

* 1. **Контрольные вопросы по дисциплине**

## Раздел 1

1. Этапы процесса Data Science.
2. Экосистема больших данных и Data Science.
3. Инфраструктуры машинного обучения
4. Перечислите типы машинного обучения.
5. Дайте определение машинного обучения с учителем, перечислите и дайте содержательную интерпретацию задачам, решаемым с помощью такого обучения.
6. Дайте определение машинного обучения без учителя, перечислите и дайте содержательную интерпретацию задачам, решаемым с помощью такого обучения.
7. Дайте определение машинного обучения с подкреплением, перечислите и дайте содержательную интерпретацию задачам, решаемым с помощью такого обучения.
8. Перечислите основные этапы решения задачи методами машинного обучения.

## Раздел 2

1. Классификация данных.
2. Работа с категориальными данными.
3. Нормирование значений признаков.
4. Задача отбора содержательных признаков и варианты ее решения.
5. Задача восстановления пропущенных данных.

## Раздел 3

1. Сформулируйте постановку задачи регрессии в ML, укажите используемые функционалы ошибки.
2. Сформулируйте постановку задачи кластеризации в ML, укажите используемые функционалы ошибки.
3. Линейная модель в задачах регрессии и классификации.
4. Градиентный метод и стохастический градиентный метод.
5. Спрямляющие пространства.
6. Проблема переобучения и подходы к ее решению.
7. Методы опорных векторов.
8. Деревья принятия решений.
9. Техники деления выборки: отложенная выборка и кросс-валидация.
10. Кривая обучения.
11. Проверочная кривая.
12. Метод сеточного поиска для настройки гиперпараметров.
13. Метрики качества задачи классификации.
14. Определение композиции алгоритмов. Правила построения.
15. Бутстреп. Бэггинг.
16. Случайный лес.
17. Бустинг.

## Раздел 4

1. Кластерный анализ, обзор проблемы.
2. Метод K средних.
3. EM-алгоритм.
4. Агломеративная иерархическая кластеризация.
5. Графовая кластеризация.
6. Оценка качества решения задачи кластеризации.
7. Задача обнаружения аномалий.
8. Параметрическое и непараметрическое восстановление плотности.
9. Задача визуализации.
10. Многомерное шкалирование.
11. Задача тематического моделирования.
12. Базовые тематические модели.
13. Регуляризация тематической модели.
14. Внутренние критерии качества тематических моделей.
15. Внешние критерии качества тематических моделей.

## Раздел 5

1. Обучение с подкреплением.
2. Элементы обучения с подкреплением.
3. Ограничения и круг вопросов.
4. Игра в крестики-нолики.

## Раздел 6

## Системы искусственного интеллекта и человек.

## Прикладной искусственный интеллект.

## Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Таблица 9.1

|  |  |  |
| --- | --- | --- |
| № п/п | Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся | Перечень основного оборудования |
| 1 | Специализированные компьютерные классы для проведения практических и лабораторных работ | персональные компьютер (для преподавателя), с доступом к Интернету,  * экран с электроприводом, * доска, * парты, * учебные (рабочие) места (ПК с доступом в Интернет) |
| 2 | Медиацентр (помещение для самостоятельной работы) | * парты, * учебные (рабочие) места (ПК с доступом в Интернет). |

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

**10.1 Перечень программного обеспечения**

Таблица 10.1

|  |  |  |
| --- | --- | --- |
| № п/п | Наименование ПО | Реквизиты договора (дата, номер, срок действия) /Режим доступа\* |
| 1 | ОС MS Windows 10 | Сублицензионный договор от 05.02.2019 г. № ЭА 18223-187  срок действия - февраль 2022 г.  Подписка Microsoft Imagine Premium  ГПД БУ от 31.01.2018 г. № Д18/223/100/036  срок действия – до 01. |
| 2 |  | Microsoft Office Professional Plus 2016 Acdmc  ГПД БУ от 01.11.2017 г. № ЭА-1744090  срок действия – бессрочный |
| 3 | Предоставление неисключительного права использования программ для ЭВМ ООО «Мираполис» | срок действия 01.01.2021 - 31.12.2021 |
| 4 | Anaconda | Свободно распространяемое ПО.  Режим доступа:  https://www.anaconda.com/products/individual#windows |
| 5 | TensorFlow | Свободно распространяемое ПО.  Режим доступа:  https://www.tensorflow.org/install |

* 1. **Литература**

## Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : руководство / С. Рашка ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2017. — 418 с. — ISBN 978-5-97060-409-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/100905 (дата обращения: 23.02.2021). — Режим доступа: для авториз. пользователей.

## Гаврилова, И.В. Основы искусственного интеллекта [Электронный ресурc]: учеб. пособие / И.В. Гаврилова, О.Е. Масленникова. — Электрон. дан. – Москва : ФЛИНТА, 2013. – 282 с. ЭБС Лань

## Боровская Е.В. Основы искусственного интеллекта [Электронный ресурc] / Е.В. Боровская, Н.А. Давыдова. – Москва: Лаборатория знаний (ранее БИНОМ. Лаборатория знаний), 2016.- 127 с. - ЭБС Лань

## Периодические издания

При изучении данной дисциплины использование периодической литературы не планируется.

## Перечень профессиональных баз данных и информационных справочных систем

1. Научная электронная библиотека elibrary.ru
2. ЭБС «Арбуз»
3. ЭБС Лань

Согласованно:

Библиотека ОмГТУ

штамп КО и подпись зам. директора библиотеки)

«\_\_\_\_\_\_»\_\_\_\_\_\_\_\_\_2020 г.