**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Поиск информации в неструктурированных данных

Information Retrieval in Unstructured Data

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 3

Регистрационный номер рабочей программы: 002230

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

Курс посвящен основам работы с данными в машинном обучении и при построении искусственного интеллекта.

После освоения курса обучающиеся должны уметь самостоятельно ориентироваться в научной литературе по машинному обучению, иметь практический опыт подготовки сырых данных, обучение моделей и применение их на практике.

Для достижения этих целей занятия разделены на три части:

1. Вводные лекции по основам машинного обучения

2. Практическое задание по работе с данными и обучение модели

3. Доклад на семинаре о результатах работы и принципах работы выбранного алгоритма.

**1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)**

Дисциплина изучается в 5-м семестре обучения. Обучающиеся должны обладать общими знаниями теории вероятности, машинного обучения и быть способными понимать научные статьи, написанные на английском языке.

**1.3. Перечень результатов обучения (learning outcomes)**

После освоения курса обучающиеся должны быть способны пройти все этапы построения системы машинного обучения и внятно описывать ее.

Дисциплина способствует развитию следующих компетенций:

* ОПК-1 — способен применять фундаментальные знания, полученные в области математических и (или) естественных наук, и использовать их в профессиональной деятельности;
* ОПК-3 — способен применять современные информационные технологии, в том числе отечественные, при создании программных продуктов и программных комплексов различного назначения;
* ОПК-4 — способен участвовать в разработке технической документации программных продуктов и программных комплексов;
* ОПК-5 — способен инсталлировать и сопровождать программное обеспечение для информационных систем и баз данных, в том числе отечественного производства;
* ПКА-1 — способен демонстрировать базовые знания математических и естественных наук, программирования и информационных технологий;
* ПКП-1 — способность проводить под научным руководством исследование на основе существующих методов в конкретной области профессиональной деятельности;
* ПКП-2 — способен решать задачи в области развития науки, техники и технологии с учетом нормативного правового регулирования в сфере интеллектуальной собственности;
* ПКП-4 — способен применять современные информационные технологии при проектировании, реализации, оценке качества и анализа эффективности программного обеспечения для решения задач в различных предметных областях;
* ПКП-5 — способен использовать основные методы и средства автоматизации проектирования, реализации, испытаний и оценки качества при создании конкурентоспособного программного продукта и программных комплексов, а также способен использовать методы и средства автоматизации, связанные с сопровождением, администрированием и модернизацией программных продуктов и программных комплексов;
* ПКП-6 — способен использовать знания направлений развития компьютеров с традиционной (нетрадиционной) архитектурой; современных системных программных средств: операционных систем, операционных и сетевых оболочек, сервисных программ; тенденции развития функций и архитектур проблемно-ориентированных программных систем и комплексов в профессиональной деятельности;
* ПКП-8 — способен использовать современные методы разработки и реализации конкретных алгоритмов математических моделей на базе языков программирования и пакетов прикладных программ моделирования;
* УКБ-3 — способен понимать сущность и значение информации в развитии общества, использовать основные методы получения и работы с информацией с учетом современных технологий цифровой экономики и информационной безопасности.

**1.4. Перечень и объём активных и интерактивных форм учебных занятий**

Активные и интерактивные формы занятий (общим объёмом 30 ак. часов): семинары, предполагают участие в семинаре в качестве докладчика и слушателя.

**Раздел 2. Организация, структура и содержание учебных занятий**

**2.1. Организация учебных занятий**

**2.1.1 Основной курс**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
| Код модуля в составе дисциплины,  практики и т.п. | Контактная работа обучающихся с преподавателем | | | | | | | | | | | | Самостоятельная работа | | | | Объём активных и интерактивных  форм учебных занятий | Трудоёмкость |
| лекции | семинары | консультации | практические  занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная  аттестация | итоговая аттестация | под руководством преподавателя | в присутствии  преподавателя | сам. раб. с использованием  методических материалов | текущий контроль (сам.раб.) | промежуточная аттестация (сам.раб.) | итоговая аттестация  (сам.раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| Форма обучения: очная | | | | | | | | | | | | | | | | | | |
| Семестр 5 |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 73 |  | 3 |  | 30 | 3 |
|  |  | 1-50 |  |  |  |  |  |  | 1-50 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО |  | 30 |  |  |  |  |  |  | 2 |  |  |  | 73 |  | 3 |  |  | 3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации | | | | | | |
| Код модуля в составе дисциплины, практики и т.п. | Формы текущего контроля успеваемости | | Виды промежуточной аттестации | | Виды итоговой аттестации  (только для программ итоговой аттестации и дополнительных образовательных программ) | |
| Формы | Сроки | Виды | Сроки | Виды | Сроки |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | |
| Форма обучения: очная | | | | | | |
| Семестр 5 |  |  | зачёт, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Часть 1. Введение в Машинное обучение

1. Виды алгоритмов и решаемых задач
   * 1. Классификация
     2. Кластеризация
     3. Временные ряды
2. Оценка качества работы алгоритмов
3. Глубокое машинное обучения и системы распознавания обзоров
4. Обучение с подкреплением

Часть 2. Практическая работа

1. Участники выбирают задачу, которую они готовы решить в рамках семинара
2. С участием преподавателя подбирается способ получить исходные данные и выбрать подходы к обучению

Часть 3. Доклады участников о проделанной работе и выбранном алгоритме.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Обучение проводится последовательно путем семинарских занятий с закреплением полученных знаний в ходе выполнения проекта и доклада о нем. На семинарах рассматриваются основные, имеющие принципиальное значение и наиболее трудные для понимания и усвоения вопросы.

**3.1.2 Методическое обеспечение самостоятельной работы**

При выполнении самостоятельной работы обучающиеся имеют возможность пользоваться специализированными источниками, приведенными в разделе 3.4.

**3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания**

Показателем, характеризующим текущую учебную работу, является вовлечённость в работу на занятиях и состоит, например, в уточняющих вопросах о своем проекте и о докладах других участников. Текущая работа обучающегося оценивается по шкале от 0 до 20 баллов.

Выполненный на семинаре доклад оценивается по шкале от 0 (нет доклада или очень плохой доклад) до 20 баллов (доклад, подробно раскрывающий тему, снабжённый качественными презентационными материалами, грамотно и интересно изложенный).

В ходе промежуточной аттестации обучающийся показывает работающий прототип системы, использованные научные статьи и набор данных. Докладывает об алгоритме и процессе разработки. На доклад и демонстрацию отводится 10 минут, после чего преподаватель вправе задавать уточняющие вопросы. Работающий прототип и доклад по прототипу оцениваются вместе, по шкале от 0 до 60 баллов.

Далее полученные таким образом баллы складываются и переводятся в оценку по следующему правилу:

|  |  |  |
| --- | --- | --- |
| Балл | Оценка ECTS | Оценка при проведении зачёта |
| от 91 до 100 | A | зачтено |
| от 81 до 90 | B |
| от 71 до 80 | C |
| от 61 до 70 | D |
| от 51 до 60 | E |
| менее 50 | F | не зачтено |

**3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)**

Предлагаются доклады по актуальным научным статьям по темам курса:

1. Классификация
2. Кластеризация
3. Временные ряды
4. Оценка качества работы алгоритмов
5. Глубокое машинное обучения и системы распознавания обзоров
6. Обучение с подкреплением

**3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса**

Оценка обучающимися содержания и качества учебного процесса по дисциплине осуществляется в установленном в СПбГУ порядке.

**3.2. Кадровое обеспечение**

**3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий**

К преподаванию дисциплины могут быть допущены преподаватели, имеющие диплом о высшем образовании или учёную степень, соответствующие профилю преподаваемой дисциплины.

**3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом**

Специальных требований нет.

**3.3. Материально-техническое обеспечение**

**3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий**

В аудитории, где проводятся занятия, необходимо наличие досок и желательно наличие средств письма на них.

**3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования**

Требуется стандартная лекционная аудитория с меловыми или маркерными досками.

**3.3.3 Характеристики специализированного оборудования**

Специальных требований нет.

**3.3.4 Характеристики специализированного программного обеспечения**

Специальных требований нет.

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Мел, маркеры, тряпки для стирания мела, губки для стирания маркерной записи.

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

1. Машинное обучение Хенрик Бринк, Джозеф Ричардс, Марк Феверолф, Питер 2017.

2. Введение в статистическое обучение с примерами на языке R Г. Джеймс, ДМК Прес 2017.

**3.4.2 Список дополнительной литературы**

Не требуется.

**3.4.3 Перечень иных информационных источников**

1. <https://www.kaggle.com/>

2. <https://opencv.org>

3. <https://gym.openai.com>

4. <https://www.tensorflow.org>

5. http://www.machinelearning.ru

**Раздел 4. Разработчики программы**

Алиев Артем Александрович, старший преподаватель каф. информатики. a.aliev@spbu.ru.