

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 27.07.2022 11:31:08
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Безопасность и этика искус-
ственного интеллекта»



СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования
**«Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)»**

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«МАШИННОЕ ОБУЧЕНИЕ»

для подготовки магистров

по направлению

09.04.01 «Информатика и вычислительная техника»

по программе

«Безопасность и этика искусственного интеллекта»

Санкт-Петербург

2021

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

д.ф.-м.н. профессор А.Н. Шевляков

Рабочая программа рассмотрена и одобрена на заседании кафедры ВТ
02.09.2021, протокол № 6

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 16.09.2021, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ИС
Общая трудоемкость (ЗЕТ)	5
Курс	1
Семестр	1
Виды занятий	
Лекции (академ. часов)	17
Практические занятия (академ. часов)	17
Все контактные часы (академ. часов)	34
Самостоятельная работа, включая часы на контроль (академ. часов)	146
Всего (академ. часов)	180
Вид промежуточной аттестации	
Экзамен (курс)	1

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«МАШИННОЕ ОБУЧЕНИЕ»

Содержание дисциплины включает в себя изучение основных моделей машинного обучения: методов поиска выбросов, метрических и логических методов обработки данных (метод ближайшего соседа, деревья решений, линейная регрессия и т.д.). Практические занятия ориентированы на исследование методов обработки данных и оценке качества построенных моделей.

SUBJECT SUMMARY

«MACHINE LEARNING»

The content of the discipline includes the study of: basic machine learning models: outlier search methods, metric and logical data processing methods (nearest neighbor method, decision trees, linear regression, etc.). The laboratory workshop is focused on researching data processing methods and assessing the quality of the constructed models.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Изучение моделей и классических алгоритмов машинного обучения (линейные, метрические, логические и байесовские модели) и формирование навыков их использования в профессиональной деятельности, и умения руководить разработкой архитектуры комплексных систем искусственного интеллекта.
2. Изучение методов построения классических моделей машинного обучения (линейные, метрические, логические и байесовские модели), оценивание их точности и адекватности.
3. Приобретение знаний в области построения и анализа моделей машинного обучения, разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий.
4. Приобретение умений в построении и анализе моделей машинного обучения различных типов (метрические, логические и вероятностные модели). Получение умений в области оценке адекватности построенных моделей. Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения поставленных задач.
5. Приобретение навыков в построении и анализе моделей машинного обучения различных типов (метрические, логические и вероятностные модели). Приобретение навыков в области оценки адекватности построенных моделей. Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе знаний, полученных при освоении программы бакалавриата или специалитета.

и обеспечивает изучение последующих дисциплин:

1. «Машинное обучение в приложениях биометрии»
2. «Доверенный искусственный интеллект»

3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

Код компетенции/ индикатора компетенции	Наименование компетенции/индикатора компетенции
ПК-26	Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач
<i>ПК-26.1</i>	<i>Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области</i>
<i>ПК-26.2</i>	<i>Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области</i>
<i>ПК-26.3</i>	<i>Разрабатывает унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий</i>
ПК-27	Способен руководить проектами по созданию комплексных систем искусственного интеллекта
<i>ПК-27.1</i>	<i>Руководит разработкой архитектуры комплексных систем искусственного интеллекта</i>

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Содержание разделов дисциплины

4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	СР, ач
1	Введение в машинное обучение и основные понятия статистики. Первичная обработка данных	4	4	23
2	Кластеризация	2	2	23
3	Линейная регрессия и ее обобщения	3	3	27
4	Задача классификации	4	4	27
5	Вероятностные алгоритмы, байесовские классификаторы	2	2	23
6	Отбор оптимального числа признаков и ансамбли алгоритмов	2	2	23
	Итого, ач	17	17	146
	Из них ач на контроль	0	0	35
	Общая трудоемкость освоения, ач/зе	180/5		

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение в машинное обучение и основные понятия статистики. Первичная обработка данных	Предмет дисциплины, её объём, содержание и связь с другими дисциплинами учебного плана. Обзор литературы по курсу. Основные понятия математической статистики. Методы обработки данных: алгоритмы поиска выборок (критерий Шовене), алгоритмы восстановления пропущенных данных (метрические и статистические методы).
2	Кластеризация	Постановка задачи кластеризации, обзор алгоритмов кластеризации (FOREL, k-means...).
3	Линейная регрессия и ее обобщения	Постановка задачи регрессии, модель линейной регрессии и ее обобщения (полиномиальная регрессия, лассо и регуляризация). Контроль качества модели для задачи регрессии.
4	Задача классификации	Постановка задачи классификации. Модели машинного обучения для задачи классификации: метод ближайшего соседа, деревья решений, линейные классификаторы, метода опорных векторов. Контроль качества модели для задачи классификации.
5	Вероятностные алгоритмы, байесовские классификаторы	Понятие об алгоритмах, выдающих вероятности. Байесовский классификатор. Контроль качества модели, выдающей вероятности.

№ п/п	Наименование темы дисциплины	Содержание
6	Отбор оптимального числа признаков и ансамбли алгоритмов	Важность задачи по отбору признаков для модели искусственного интеллекта. Методы отбора признаков: прямые и итерационные. Синтез новых признаков. Понятие об ансамбле алгоритмов. Виды ансамблей для задач классификации и регрессии. Градиентный бустинг.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Введение в машинное обучение и основные понятия статистики. Первичная обработка данных	4
2. Кластеризация	2
3. Линейная регрессия и ее обобщения	3
4. Задача классификации	4
5. Вероятностные алгоритмы, байесовские классификаторы	2
6. Отбор оптимального числа признаков и ансамбли алгоритмов	2
Итого	17

4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

4.5 Реферат

Реферат не предусмотрен.

4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.

4.7 Доклад

Доклад не предусмотрен.

4.8 Кейс

Кейс не предусмотрен.

4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Материалы онлайн-курса «Машинное обучение» (видео лекций и примеры задач) размещены по адресу: <https://stepik.org/course/8057>.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

Самостоятельное изучение студентами теоретических основ дисциплины обеспечено необходимыми учебно методическими материалами (учебники, онлайн-версия курса), выполненными в печатном или электронном виде.

По каждой теме содержания рабочей программы могут быть предусмотрены индивидуальные домашние задания (расчетно-графические работы, рефераты, конспекты изученного материала, доклады и т.п.).

Изучение студентами дисциплины сопровождается проведением регулярных консультаций преподавателей, обеспечивающих практические занятия по дисциплине, за счет бюджета времени, отводимого на консультации (внеаудиторные занятия, относящиеся к разделу «Самостоятельные часы для изучения дисциплины»).

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	48
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	0
Выполнение домашних заданий, домашних контрольных работ	32
Подготовка к лабораторным работам, к практическим и семинарским занятиям	31
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференцированному зачету, экзамену	35
ИТОГО СРС	146

5 Учебно-методическое обеспечение дисциплины

5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библи.
Основная литература		
1	Рассел, Стюарт. Искусственный интеллект: современный подход [Текст] : пер. с англ. / С. Рассел, П. Норвиг, 2007. -1407 с.	45
2	Частиков А.П. Разработка экспертных систем. Среда CLIPS [Текст] : Учеб. пособие / А.П.Частиков, Т.А.Гаврилова, Д.Л.Белов, 2003. -606 с.	81
Дополнительная литература		
1	Джексон, Питер. Введение в экспертные системы [Текст] : [Учеб. пособие] / П. Джексон; Пер. с англ. и ред. В.Т.Тертышного, 2001. -622 с.	18

5.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	Портал: Искусственный интеллект. – URL: http://ru.wikipedia.org/wiki/Портал:Искусственный интеллект
2	Курс лекций по дисциплине «Системы искусственного интеллекта». – URL: http://www.mari.ru/mmlab/home/AI/
3	Онлайн-курс Шевлякова А.Н. «Машинное обучение» – URL: http://www.stepik.org/course/8057

5.3 Адрес сайта курса

Адрес сайта курса: stepik.org/course/8057

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Машинное обучение» формой промежуточной аттестации является экзамен.

Экзамен

Оценка	Описание
Неудовлетворительно	Студент продемонстрировал существенные пробелы в знаниях основного учебного материала, допустил принципиальные ошибки в выполнении предусмотренных программой заданий.
Удовлетворительно	Студент продемонстрировал знание основного учебного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии, справился с выполнением заданий, предусмотренных программой, обладает необходимыми знаниями, но допустил неточности в ответах на аттестационном испытании и при выполнении учебных заданий.
Хорошо	Студент продемонстрировал полное знание учебного материала, успешно выполнил предусмотренные программой задачи, освоил основную рекомендованную литературу, показал систематический характер знаний по дисциплине и способен к их самостоятельному пополнению и обновлению в ходе дальнейшей учебы и профессиональной деятельности.
Отлично	Студент продемонстрировал всестороннее систематическое знание учебного материала, умение свободно выполнять практические задания, освоил основную литературу и ознакомился с дополнительной литературой, рекомендованной рабочей программой дисциплины, усвоил взаимосвязь основных понятий дисциплины в их значении для приобретаемой профессии, проявил творческие способности в понимании, изложении и использовании учебного материала.

Особенности допуска

Студенты допускаются на экзамен при условии посещения ими лекционных и практических занятий (не менее 80%). Для допуска к экзамену требуется набрать проходной балл не менее 60% при прохождении теста, иметь не менее чем удовлетворительную оценку по результатам выполнения контрольной работы и коллоквиума.

6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

Примерные вопросы к экзамену

№ п/п	Описание
1	Восстановление пропущенных данных с помощью коэффициента корреляции
2	Критерий Шовене в задаче поиска выбросов
3	Описание построения и работы модели линейной регрессии
4	Описание алгоритма наивного Байеса и контроль качества его работы
5	Описание градиентного бустинга на примере деревьев небольшой высоты
6	Описание генетического алгоритма в задаче по отбору признаков

Вариант экзаменационного теста

Экзаменационный тест содержит 6 вопросов из разных тем курса. Проходной балл теста 70%.

1. Вопрос

Первая и третья квартиль значений признака P равны 2, 4 соответственно. Какие из следующих значений будут считаться выбросами?

Варианты ответа:

-1,5

0,5

0

3

6,5

7,5

8

2. Вопрос

Набор объектов A,B,C,D,E,F,G планируется разбить на 3 кластера. Для этого был применен графовый алгоритм и было построено остовное дерево минимального веса, состоящее из ребер AE, BD, CD, DF, ED, EG. Длины ребер соответственно равны 1, 2, 4, 1, 3, 2. Какие объекты попадут в один кластер с объектом A?

Варианты ответа:

D

C

B

G

E

F

3. Вопрос

Какие существуют проблемы, связанные с построением модели линейной регрессии?

Варианты ответа:

мультиколлинеарность

бесконечное множество решений у системы уравнений, которая возникает при нахождении весов модели

большие по модулю значения весов

4. Вопрос

Искусственный нейрон имеет 2 входа с весами 1, 2 соответственно. Ко входному сигналу прибавляется число 1 и применяется сигмоидная функция. Чему будет равно значение на выходе из нейрона, если на первый вход было подано число 1, а на второй вход было подано число -1?

Варианты ответа:

1

-0,5

0

0,5

5. Вопрос

Где здесь формула Байеса?

Варианты ответа:

$$\Pr(B|A)=\Pr(A|B)/\Pr(B)*\Pr(A)$$

$$\Pr(B|A)=\Pr(A|B)*\Pr(A)/\Pr(B)$$

$$\Pr(B|A)=\Pr(A|B)*\Pr(B)/\Pr(A)$$

$$\Pr(B|A)=\Pr(A|B*B/A)$$

6. Вопрос

Вспомните работу алгоритма Random Forest (случайный лес). К какому типу ансамбля он относится?

Варианты ответа:

градиентный бустинг

взвешенное голосование

голосование по большинству (комитет)

Варианты тестовых заданий для экзамена даны в материалах онлайн-курса Шевлякова А.Н. «Машинное обучение» - URL: <http://www.stepik.org/course/8057>.

Образцы задач (заданий) для контрольных (проверочных) работ

Контрольная работа состоит из 5 заданий. Она может быть проведена как в форме компьютерного тестирования с автоматизированным оцениванием, так и в бумажном виде.

Образец задачи (задания)

Однажды я попросил, чтобы студенты ответили на два вопроса анкеты «ваш год рождения» и «ваш возраст».

Из их ответов я сформировал таблицу, в которой был столбец P = «год рождения студента» и Q = «возраст студента».

Оказывается, значение коэффициента корреляции признаков P и Q зависит от месяца, в котором проводилось анкетирование (это не шутка!). Укажите два месяца, которым соответствует наименьшее (по модулю) значение коэффициента корреляции признаков P и Q .

Варианты ответов: январь, февраль, март, апрель, май, июнь, июль, август, сентябрь, октябрь, ноябрь, декабрь.

Образцы задач (заданий) для контрольных (проверочных) работ даны в материалах онлайн-курса Шевлякова А.Н. «Машинное обучение» - URL: <http://www.>

Тест состоит из 10 вопросов

Вопрос (пример)

Искусственный нейрон имеет 2 входа с весами 1, 2 соответственно. Ко входному сигналу прибавляется число 1 и применяется сигмоидная функция. Чему будет равно значение на выходе из нейрона, если на первый вход было подано число 1, а на второй вход было подано число -1?

Варианты ответа:

1

-0,5

0

0,5

Вопросы коллоквиума по теме ”Введение в машинное обучение и основные понятия статистики. Первичная обработка данных”

- 1) Предмет дисциплины Машинное обучение
- 2) Основные понятия математической статистики.
- 3) Методы обработки данных
- 4) Алгоритмы поиска выбросов (критерий Шовене),
- 5) Алгоритмы восстановления пропущенных данных (метрические и статистические методы).

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
4	Введение в машинное обучение и основные понятия статистики. Первичная обработка данных	
5		Коллоквиум
12	Задача классификации	
13		Тест
15	Отбор оптимального числа признаков и ансамбли алгоритмов	
16		Контрольная работа

6.4 Методика текущего контроля

на лекционных занятиях

Текущий контроль включает в себя контроль посещаемости (не менее 80% занятий).

на практических (семинарских) занятиях

Текущий контроль включает в себя контроль посещаемости (не менее 80% занятий).

В ходе проведения практических занятий в форме коллоквиума целесообразно привлечение студентов к как можно более активному участию в дискуссиях, решении задач, обсуждениях и т. д. При этом активность студентов также может учитываться преподавателем, как один из способов текущего контроля на практических занятиях. Участие в **коллоквиуме** оценивается по системе зачтено/ не зачтено:

«не зачтено» - ставится, если основное содержание вопроса не раскрыто, не даны ответы на вопросы преподавателя, допущены грубые ошибки в определении понятий и в использовании терминологии;

«зачтено» ставится, если продемонстрировано усвоение основного содержания материала, ответы на вопросы даны полностью.

Текущий контроль в форме **тестирования** включает в себя прохождение теста из 10 вопросов с проходным баллом 60%. Для получения оценки «удовле-

творительно» необходимо дать правильные ответы на не менее 60% тестовых вопросов, «хорошо» - не менее 75%, «отлично» - на 90% или более.

Промежуточный контроль в виде **контрольной работы** может быть проведен как в форме компьютерного тестирования с автоматизированным оцениванием заданий, так и в бумажном виде. Контрольная работа содержит 5 заданий. Для получения оценки «удовлетворительно» необходимо правильно выполнить 2 задания, «хорошо» - 4 задания, «отлично» - 5 заданий.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на лекционных и практических занятиях студентов по методикам, описанным выше.

7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	1) Количество посадочных мест – в соответствии с контингентом, 2) рабочее место преподавателя, персональный компьютер IBM, совместимый Pentium или выше, проектор, экран/интерактивная панель, меловая/маркерная доска.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Практические занятия	Аудитория	1) Количество посадочных мест – в соответствии с контингентом, 2) рабочее место преподавателя, персональный компьютер IBM, совместимый Pentium или выше, проектор, экран/интерактивная панель, меловая/маркерная доска.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше

8 Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Дата	Изменение	Дата и номер протокола заседания УМК	Автор	Начальник ОМОЛА