

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 26.06.2023 14:15:20
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Информационно-управляющие
системы»



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 бакалавров

по направлению

09.03.02 «Информационные системы и технологии»

по профилю

«Информационно-управляющие системы»

Санкт-Петербург

2022

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

доцент, к.т.н., доцент Новикова Е.С.

Рабочая программа рассмотрена и одобрена на заседании кафедры ИС
21.02.2022, протокол № 2

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 24.02.2022, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ИС
Общая трудоемкость (ЗЕТ)	4
Курс	3
Семестр	5
Виды занятий	
Лекции (академ. часов)	17
Практические занятия (академ. часов)	17
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	35
Самостоятельная работа, включая часы на контроль (академ. часов)	109
Всего (академ. часов)	144
Вид промежуточной аттестации	
Дифф. зачет (семестр)	5

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«МАШИННОЕ ОБУЧЕНИЕ»

Данный курс знакомит слушателей с основными методами анализа данных. Рассматриваются основные этапы интеллектуального анализа, включая подготовку исходных данных, особенности выбора параметров алгоритмов, оценку и представление результатов анализа. Изучаются основные задачи интеллектуального анализа данных: классификация, кластеризация, поиск ассоциативных правил. На практике разбираются основные алгоритмы решения этих задач и особенности их применения для больших данных.

Данный курс позволяет детально изучить понятия, задачи и методы интеллектуального анализа данных, предназначенные для создания математических моделей, адекватных окружающему миру и решения прикладных задач.

SUBJECT SUMMARY

«MACHINE LEARNING»

This course introduces students to the main methods of data analysis. The main stages of data mining , including the source data preparation, the setting up of the initial parameters of the algorithms, evaluation and presentation of the analysis results are considered. The main tasks of data mining are studied: classification, clustering, search for associative rules. In practice, the basic algorithms for solving these problems and the peculiarities of their application for big data are analyzed.

Course "Machine Learning" introduces students to the concepts, problems and techniques of data mining for creating mathematical models adequate to the world and solution of the practical tasks.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Целью освоения дисциплины «Машинное обучение» являются приобретение студентами теоретических знаний и формирование практических умений и навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения для решения практических задач по анализу данных.

2. Задачами изучения дисциплины являются:

-изучение содержательной и математической постановки основных задач машинного обучения, получение знаний о методах машинного обучения и предобработки исходных данных;

-формирование знаний об основных понятиях, этапах анализа и интерпретации данных;

-формирование умений по выбору наиболее подходящего способа решения задачи и обоснованию его эффективности;

-формирование навыков по представлению результатов работы алгоритмов машинного обучения;

-освоение практических навыков по сбору данных, методов их анализа и обобщения, выбору средств и методов обработки данных для решения практических задач; представлению результатов анализа в форме, поддерживающей принятие решения в соответствии с поставленной задачей.

3. Студенты получают знания по ключевым понятиям, целям и задачам использования машинного обучения для решения различных практических задач; изучают методологические основы применения алгоритмов машинного обучения.

4. Дисциплина формирует умения по выбору метода машинного обучения, соответствующего научно-практической задаче, и его обоснованию, умения по

объяснению и визуализации результатов работы алгоритмов машинного обучения, а также оценке качества моделей анализа.

5. Дисциплина формирует практические навыки по применению программных инструментов и библиотек по предобработке исходных данных, построению моделей анализа, оценки качества моделей и визуализации полученных результатов.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Архитектура информационных систем»
2. «Теория вероятностей и математическая статистика»

и обеспечивает изучение последующих дисциплин:

1. «Большие данные»
2. «Компьютерная математика»
3. «Методы искусственного интеллекта»

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
ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решения задач профессиональной деятельности;
<i>ОПК-2.1</i>	<i>Знает современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности</i>
<i>ОПК-2.2</i>	<i>Умеет выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности</i>
<i>ОПК-2.3</i>	<i>Имеет навыки применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности</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	ИКР, ач	СР, ач
1	Введение в машинное обучение.	1		0	7
2	Задача классификации. Линейные классификаторы.	3	4		30
3	Задача классификации. Деревья решений.	4	5		20
4	Задача классификации и регрессионный анализ.	2	2		15
5	Ансамблевые методы.	2			15
6	Задача кластеризации.	4	6		20
7	Заключение	1		1	2
	Итого, ач	17	17	1	109
	Из них ач на контроль	0	0	0	0
	Общая трудоемкость освоения, ач/зе	144/4			

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение в машинное обучение.	Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Основные этапы эволюции алгоритмов машинного обучения.
2	Задача классификации. Линейные классификаторы.	Линейные классификаторы. Общий вид метрического классификатора. Алгоритм К ближайших соседей. Метод опорных векторов. Линейные перцептрон
3	Задача классификации. Деревья решений.	Деревья решений. Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев.
4	Задача классификации и регрессионный анализ.	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия.
5	Ансамблевые методы.	Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг

№ п/п	Наименование темы дисциплины	Содержание
6	Задача кластеризации.	Формальная постановка задачи кластеризации. Алгоритмы кластеризации с фиксированным количеством кластеров. Алгоритмы кластеризации по плотности. Иерархическая кластеризация.
7	Заключение	

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Методы классификации. Алгоритм К ближайших соседей.	2
2. Методы классификации. Метод опорных векторов	2
3. Методы классификации. Деревья решений. Правила и анализ качества (точность, полнота).	2
4. Методы классификации. Лес решающих деревьев.	3
5. Методы классификации. Линейная регрессия.	2
6. Алгоритмы кластеризации с фиксированным количеством кластеров.	2
7. Алгоритмы кластеризации по плотности.	2
8. Иерархическая кластеризация.	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.

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.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	15

Текущая СРС	Примерная трудоемкость, ач
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	5
Самостоятельное изучение разделов дисциплины	10
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	49
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	30
ИТОГО СРС	109

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	Методы оперативного и интеллектуального анализа данных [Текст] : метод. указания к лаб. работам / Санкт-Петербургский государственный электротехнический университет им. В.И. Ульянова (Ленина) "ЛЭТИ", 2008. -32 с.	85
2	Элбон, Крис. Машинное обучение с использованием Python. Сборник рецептов [Текст] : [пер. с англ.] / К. Элбон, 2020. -369 с.	20
3	Федоров, Дмитрий Юрьевич. Программирование на языке высокого уровня Python [Текст] : учеб. пособие для приклад. бакалавриата вузов по инженер.-техн. направлениям / Д. Ю. Федоров, 2018. -125, [1] с.	50
Дополнительная литература		
1	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение [Электронный ресурс] / Дж. Вандер Плас, 2018. -576 с.	неогр.
2	Бенджамин Бенгфорт Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка [Электронный ресурс] / Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони, 2021. -368 с.	неогр.

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

№ п/п	Электронный адрес
1	Е. Соколов. Курс лекций по машинному обучению ВШЭ. https://github.com/esokolov

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=6930>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Машинное обучение» предусмотрены следующие формы промежуточной аттестации: зачет с оценкой.

Зачет с оценкой

Оценка	Описание
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок
Хорошо	Студент овладел курсом, но в отдельных вопросах испытывает затруднения. Умеет решать задачи
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач

Особенности допуска

Для допуска к дифференцированному зачету студент должен посетить не менее 80% занятий, выполнить и сдать все практические работы в срок. При сдаче практических работ допускается задержка на 1 неделю.

В ходе дифференцированного зачета студент должен ответить на два теоретических вопроса по курсу, при не точных или неудовлетворительных ответах, студенту могут быть заданы дополнительные вопросы.

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

Вопросы к дифф.зачету

№ п/п	Описание
1	Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
2	Предобработка данных. Масштабирование. Нормирование и стандартизация.
3	Предобработка данных. Полиномиальные признаки. One-hot encoding.
4	Кластеризация. kMeans, DBSCAN.
5	Алгоритмы и параметры иерархической кластеризацией. Интерпретация дендрограмм
6	Алгоритмы кластеризации по плотности.
7	Алгоритмы кластеризации с фиксированным количеством кластеров.
8	Оценка качества кластеризации. Внешние и внутренние метрики.
9	Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
10	Ансамблевые методы. Bagging. Случайные леса. AdaBoost.
11	Бустинг деревьев решений.
12	Ошибка внутри и вне выборки. Валидация и кросс-валидация.
13	Линейная регрессия. Полиномиальная регрессия.
14	Логистическая регрессия. Градиентный спуск.
15	Метрики качества классификации. Precision-Recall и ROC кривые. AUC.
16	Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
17	Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев.
18	Деревья решений. Анализ с помощью ROC кривой.
19	Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм.
20	Метрические классификаторы. kNN. WkNN. Отбор эталонов. Kdtree

Форма билета

Министерство науки и высшего образования Российской Федерации
ФГАОУ ВО «Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» имени В.И. Ульянова (Ленина)»

БИЛЕТ № 1

Дисциплина **Машинное обучение** ФКТИ

1. Кластеризация. kMeans, DBSCAN.
2. Метрики качества классификации. Precision-Recall и ROC кривые. AUC.

УТВЕРЖДАЮ

Заведующий кафедрой

В.В. Цехановский

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
2	Задача классификации . Линейные классификаторы	
3		Практическая работа
4	Задача классификации . Линейные классификаторы	
5		Практическая работа
6	Задача классификации. Деревья решений.	
7		Практическая работа
8	Задача классификации. Деревья решений.	
9		Практическая работа
10	Задача классификации и регрессионный анализ.	
11		Практическая работа
12	Задача кластеризации.	
13		Практическая работа
14	Задача кластеризации.	
15		Практическая работа
16	Задача кластеризации.	
17		Практическая работа

6.4 Методика текущего контроля

Текущий контроль включает в себя контроль посещаемости на лекционных и практических занятиях (не менее **80 %** занятий), по результатам которого студент получает допуск на дифференцированный зачет.

В ходе проведения практических занятий 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. Другим критерием является полнота и своевременность выполнения практических работ.

Выполнение практических работ оценивается по четырех балльной шкале:

– «Отлично» – оцениваются практические работы, в которых точно и четко описаны исходные данные, постановка задачи, описаны и обоснованы

выбранные параметры модели анализа, полученные результаты обоснованы и продемонстрированы на рисунках. Выводы содержательны. Отчет о выполнении практической работы отвечает всем требованиям по оформлению и содержанию. Представлен программный код, реализующий практическую работу. Студент полно отвечает на два дополнительных вопроса по теме практической работе.

– «Хорошо» – оцениваются практические работы, в которых представлены исходные данные, постановка задачи, описаны и обоснованы выбранные параметры модели анализа, полученные результаты обоснованы и продемонстрированы на рисунках. Допускаются недостатки в обосновании выбора параметров модели анализа, графическом материале отчета по практической работе и сделанных выводах. Отчет о выполнении практической работы отвечает всем требованиям по оформлению и содержанию. Представлен программный код, реализующий практическую работу. Студент отвечает на два дополнительных вопроса по теме практической работе.

– «Удовлетворительно» – оцениваются практические работы, в которых частично представлены исходные данные, постановка задачи, описаны выбранные параметры модели анализа, полученные результаты частично обоснованы. Графический материал отсутствует, студент затрудняется объяснить и обосновать выбор параметров модели анализа. Выводы очень краткие. Отчет о выполнении практической работы не соответствует всем требованиям по оформлению и содержанию. Представлен программный код, реализующий практическую работу, студент отвечает на вопросы по коду программы. Студент отвечает неполно на два дополнительных вопроса по теме практической работе.

– «Неудовлетворительно» – оцениваются практические работы, в которых отсутствуют исходные данные, постановка задачи, частично описаны и обоснованы выбранные параметры модели анализа, полученные результаты не обоснованы. Графический материал отсутствует, студент затрудняется объяснить и

обосновать выбор параметров модели анализа. Отчет о выполнении практической работы не отвечает всем требованиям по оформлению и содержанию. Представлен программный код, реализующий практическую работу. Студент не отвечает на вопросы по коду программы. Ответы на два дополнительных вопроса - неудовлетворительны.

Сдача практических работ выполняется по выше представленному графику, при сдаче работы не в срок оценка за нее снижается на один балл.

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	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, проектор, экран, персональный компьютер или ноутбук	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, проектор, экран, персональный компьютер или ноутбук	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