

ИНСТИТУТ ИНТЕЛЛЕКТУАЛЬНЫХ КИБЕРНЕТИЧЕСКИХ СИСТЕМ

КАФЕДРА КИБЕРНЕТИКИ

ОДОБРЕНО УМС ИИКС

Протокол № УМС-575/01-1

от 30.08.2021 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ ЧАСТЬ 2

Направление подготовки
(специальность)

[1] 09.04.04 Программная инженерия

Семестр	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	В форме практической подготовки/В СРС, час.	КСР, час.	Форма(ы) контроля, экс./зач./КР/КП
3	5-7	180- 252	32	0	16	96- 168	0	Э
Итого	5-7	180- 252	32	0	16	0 96- 168	0	

АННОТАЦИЯ

Машинное обучение является современным подходом к созданию систем, которые в процессе работы способны вырабатывать решения не по заранее запрограммированным правилам, а на основе предварительного анализа располагаемых данных. Применение методов машинного обучения позволяет создавать интеллектуальные системы управления, распознавания образов, прогнозирования во всех сферах деятельности человека. В курсе студенты знакомятся с основами машинного обучения: постановкой задачи обучения, подготовкой данных, принципами обучения и принятия решения, подходами к организации обучения и валидации результатов, методами и алгоритмами классификации и кластеризации данных. Теоретические основы машинного обучения закрепляются на лабораторных занятиях и используются при выполнении самостоятельного исследования в рамках индивидуальной курсовой работы.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью освоения учебной дисциплины «Машинное обучение» является овладение будущими специалистами теоретическими знаниями и практическими подходами к построению систем, способных принимать адекватные интеллектуальные решения, сравнимые по качеству и точности с решениями, которые принимает человек. Необходимость создания таких систем возрастает. В то же время растет и возможность их реализации в связи с развитием технологической основы вычислительных систем и средств коммуникаций. Изучение основных задач машинного обучения, методов их решения, алгоритмического и программного обеспечения необходимо будущим специалистам в области программной инженерии при создании современных программных систем, предназначенных для интеллектуальной обработки больших объемов данных в условиях неполноты или отсутствия их формального описания.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина «Машинное обучение» относится к общенаучному циклу М1 образовательной программы и является обязательной для магистранта. Дисциплина требует от слушателя общематематической подготовки по дискретной математике, математическому анализу, линейной алгебре, дифференциальным уравнениям, методам оптимизации, а также углубленного знания теории вероятностей и математической статистики. Полученные знания могут быть применены студентами в процессе обучения при выполнении текущей научно-исследовательской и выпускной работ. Дисциплина формирует систему базовых понятий, необходимых для специалиста в области программной инженерии, прикладной математики и информатики, способствует выработке профессиональных решений практических задач.

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные и(или) общепрофессиональные компетенции:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
--------------------------------	--

Профессиональные компетенции в соответствии с задачами и объектами (областями знаний) профессиональной деятельности:

Задача профессиональной деятельности (ЗПД)	Объект или область знания	Код и наименование профессиональной компетенции; Основание (профессиональный стандарт-ПС, анализ опыта)	Код и наименование индикатора достижения профессиональной компетенции
производственно-технологический			
разработка, тестирование и сопровождение программного обеспечения, применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения, взаимодействие с заказчиком в процессе выполнения	инструменты разработки программного продукта, процессы жизненного цикла программного продукта	ПК-10.2 [1] - Способен использовать технологии, методы и инструментальные средства обработки больших данных <i>Основание:</i> Профессиональный стандарт: 06.042	З-ПК-10.2[1] - Знать технологии, методы и инструментальные средства обработки больших данных; У-ПК-10.2[1] - Уметь пользоваться методами и инструментами получения, хранения, передачи, обработки больших данных; В-ПК-10.2[1] - Владеть методами и инструментами получения, хранения, передачи, обработки больших данных
разработка, тестирование и сопровождение программного обеспечения, применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения, взаимодействие с заказчиком в процессе выполнения	инструменты разработки программного продукта, процессы жизненного цикла программного продукта	ПК-15 [1] - способен применять навыки создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов <i>Основание:</i> Профессиональный стандарт: 06.028, 06.042	З-ПК-15[1] - Знать: технологии создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов ; У-ПК-15[1] - Уметь: применять навыки создания программного обеспечения для анализа, распознавания и обработки информации, систем цифровой обработки сигналов ; В-ПК-15[1] - Владеть: навыками создания программного обеспечения для

			анализа, распознавания и обработки информации, систем цифровой обработки сигналов
освоение и применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-9.2 [1] - Способен применять методы получения, хранения, передачи и обработки больших данных <i>Основание:</i> Профессиональный стандарт: 06.042	3-ПК-9.2[1] - Знать методы получения, хранения, передачи и обработки больших данных; У-ПК-9.2[1] - Уметь использовать методы получения, хранения, передачи и обработки больших данных; В-ПК-9.2[1] - Владеть методами получения, хранения, передачи и обработки больших данных
освоение и применение методов и инструментальных средств управления инженерной деятельностью и процессами жизненного цикла программного обеспечения	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-9.3 [1] - Способен использовать методы машинного обучения <i>Основание:</i> Профессиональный стандарт: 06.042	3-ПК-9.3[1] - Знать методы машинного обучения; У-ПК-9.3[1] - Уметь методы машинного обучения; В-ПК-9.3[1] - Владеть методами машинного обучения
научно-исследовательский			
построение моделей объектов профессиональной деятельности с использованием инструментальных средств компьютерного моделирования	программный продукт, процессы, методы и инструменты разработки программного продукта	ПК-4 [1] - способен применять существующие методы и алгоритмы решения задач распознавания и обработки данных <i>Основание:</i> Профессиональный стандарт: 40.011	3-ПК-4[1] - Знать: методы и алгоритмы решения задач распознавания и обработки данных ; У-ПК-4[1] - Уметь: применять методы и алгоритмы решения задач распознавания и обработки данных ; В-ПК-4[1] - Владеть: методами и алгоритмами решения задач распознавания и обработки данных
построение моделей объектов профессиональной деятельности с использованием инструментальных	программный продукт, процессы, методы и инструменты разработки программного	ПК-5 [1] - способен применять существующие методы и алгоритмы решения задач цифровой обработки сигналов	3-ПК-5[1] - Знать: методы и алгоритмы решения задач цифровой обработки сигналов ; У-ПК-5[1] - Уметь:

средств компьютерного моделирования	продукта	Основание: Профессиональный стандарт: 40.011	применять методы и алгоритмы решения задач цифровой обработки сигналов ; В-ПК-5[1] - Владеть: методами и алгоритмами решения задач цифровой обработки сигналов
-------------------------------------	----------	---	---

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Разделы учебной дисциплины, их объем, сроки изучения и формы контроля:

№ п.п	Наименование раздела учебной дисциплины	Недели	Лекции/ Практи. (семинары) / Лабораторные работы, час.	Обязат. текущий контроль (форма*, неделя)	Максимальный балл за раздел**	Аттестация раздела (форма*, неделя)	Индикаторы освоения компетенции
	<i>3 Семестр</i>						
1	Постановка задачи машинного обучения. Препроцессинг данных. Методы классификации данных. Композиции методов классификации.	1-8		к.р-8 (30)	30	к.р-8	3-ПК-10.2, У-ПК-10.2, В-ПК-10.2, 3-ПК-15, У-ПК-15, В-ПК-15, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-5, У-ПК-5, В-ПК-5, 3-ПК-9.2,

							У-ПК-9.2, В-ПК-9.2, 3-ПК-9.3, У-ПК-9.3, В-ПК-9.3
2	Обучение без учителя. Кластеризация данных. Решение задач машинного обучения на многослойных нейронных сетях. Глубокое обучение.	9-16		к.р-16 (30)	30	к.р-16	3-ПК-10.2, У-ПК-10.2, В-ПК-10.2, 3-ПК-15, У-ПК-15, В-ПК-15, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-5, У-ПК-5, В-ПК-5, 3-ПК-9.2, У-ПК-9.2, В-ПК-9.2, 3-ПК-9.3,

							У-ПК-9.3, В-ПК-9.3
	<i>Итого за 3 Семестр</i>		32/0/16		60		
	Контрольные мероприятия за 3 Семестр				40	Э	3-ПК-10.2, У-ПК-10.2, В-ПК-10.2, 3-ПК-15, У-ПК-15, В-ПК-15, 3-ПК-4, У-ПК-4, В-ПК-4, 3-ПК-5, У-ПК-5, В-ПК-5, 3-ПК-9.2, У-ПК-9.2, В-ПК-9.2, 3-ПК-9.3, У-ПК-9.3, В-ПК-9.3

* – сокращенное наименование формы контроля

** – сумма максимальных баллов должна быть равна 100 за семестр, включая зачет и (или) экзамен

Сокращение наименований форм текущего контроля и аттестации разделов:

Обозначение	Полное наименование
к.р	Контрольная работа
Э	Экзамен

КАЛЕНДАРНЫЙ ПЛАН

Недели	Темы занятий / Содержание	Лек., час.	Пр./сем., час.	Лаб., час.
	<i>3 Семестр</i>	32	0	16
1-8	Постановка задачи машинного обучения. Препроцессинг данных. Методы классификации данных. Композиции методов классификации.	16		8
1 - 2	Постановка задачи машинного обучения. Препроцессинг данных. Метод главных компонент. Типы задач машинного обучения. Примеры прикладных задач. Основные понятия и определения. Функционал качества. Объекты и признаки. Методы предобработки исходных данных. Метод главных компонент. Проблема переобучения и понятие обобщающей способности. Методика тестирования обучаемых алгоритмов. Кросс-валидация. Приёмы генерации модельных данных.	Всего аудиторных часов		
		4		2
		Онлайн		
3 - 6	Задачи классификации данных. Логические и метрические методы. Метод опорных векторов. Постановка задачи классификации. Байесовский подход к классификации. Наивный байесовский классификатор. Нормальный дискриминантный анализ. Линейный дискриминант Фишера. Метод опорных векторов. Метод ближайших соседей. Деревья решений. Бинарная классификация. Оценка точности классификации. Специфичность и чувствительность. ROC-анализ.	Всего аудиторных часов		
		8		4
		Онлайн		
7 - 8	Композиции методов классификации. Повышение точности решения. Подходы к построению ансамблей классификаторов. Базовый классификатор и решающее правило. Бустинг, стекинг, бэггинг. Бутстреп-метод. Алгоритм AdaBoost. Случайный лес.	Всего аудиторных часов		
		4		2
		Онлайн		
9-16	Обучение без учителя. Кластеризация данных. Решение задач машинного обучения на многослойных нейронных сетях. Глубокое обучение.	16		8
9 - 11	Обучение без учителя. Кластеризация данных. Оценка валидности кластеризации. Постановка задачи кластеризации данных. Агломеративные методы кластеризации. Метод Уорда. Представление результатов иерархической кластеризации.	Всего аудиторных часов		
		6		3
		Онлайн		

	Дендрограмма. Метод k-средних. Статистический анализ качества кластеризации данных. Силуэты кластеров, индексы Данна и ДэвисаБолдина.			
12 - 14	Решение задач машинного обучения на многослойных нейронных сетях. Математическая модель многослойной нейронной сети. Многослойный персептрон. Постановка задачи обучения. Метод обратного распространения ошибки. Применение многослойного персептрона для решения задач машинного обучения с учителем. Самообучающиеся нейронные сети. Сеть Кохонена. Карта Кохонена. Особенности применения самообучающихся нейронных сетей для решения задач кластеризации данных.	Всего аудиторных часов		
		6		3
		Онлайн		
15 - 16	Новый взгляд на задачи машинного обучения. Глубокое обучение. Проблема формирования вектора признаков. Сети глубокого доверия. Метод обучения. Предобучение слоёв глубокой сети. Автоэнкодеры. Ограниченная машина Больцмана. Свёрточные нейронные сети. Математическая модель сети. Ядро свёртки. Особенности архитектуры свёрточных сетей. Области практического применения.	Всего аудиторных часов		
		4		2
		Онлайн		

Сокращенные наименования онлайн опций:

Обозначение	Полное наименование
ЭК	Электронный курс
ПМ	Полнотекстовый материал
ПЛ	Полнотекстовые лекции
ВМ	Видео-материалы
АМ	Аудио-материалы
Прз	Презентации
Т	Тесты
ЭСМ	Электронные справочные материалы
ИС	Интерактивный сайт

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

Недели	Темы занятий / Содержание
	<i>3 Семестр</i>
1 - 2	Методы предобработки исходных данных. Метод главных компонент. Методы предобработки исходных данных. Метод главных компонент.
3 - 6	Метрические и логические методы классификации. Метрические и логические методы классификации.
7 - 8	Композиции методов классификации. Композиции методов классификации.
9 - 11	Методы кластеризации данных. Методы кластеризации данных.
12 - 14	Решение задач машинного обучения на многослойных

	нейронных сетях. Решение задач машинного обучения на многослойных нейронных сетях.
15 - 16	Методы глубокого обучения. Методы глубокого обучения.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Занятия проводятся по классической схеме чтения лекций и проведения лабораторных занятий. Контроль самостоятельной работы и освоения изученной темы проводится на лабораторных занятиях. В процессе контрольного опроса (КО) каждый слушатель должен ответить на вопросы преподавателя или провести краткие расчеты по предложенным задачам. Результаты контрольного опроса оцениваются в баллах. Для усвоения методов и алгоритмов машинного обучения студенты выполняют исследовательскую курсовую работу по индивидуальному заданию. В процессе выполнения курсовой работы студенты знакомятся с дополнительной литературой по теме проекта. Преподаватель осуществляет текущий контроль результатов исследовательской работы студента и учитывает их в течение семестра в оценках за разделы курса. Курсовая работа студента, представленная в форме Пояснительной записки, содержит описание всех этапов выполнения задания и анализ полученных результатов. Самостоятельная работа студента включает: 1) ознакомление с литературой по темам курса, изучение электронных учебников или курсов по машинному обучению и анализу данных; 2) подготовку к лабораторным занятиям; 3) изучение прикладных программ пакетов, используемых в лабораторном практикуме.

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущего, рубежного и промежуточного контроля по дисциплине.

Связь между формируемыми компетенциями и формами контроля их освоения представлена в следующей таблице:

Компетенция	Индикаторы освоения	Аттестационное мероприятие (КП 1)
ПК-10.2	З-ПК-10.2	Э, к.р-8, к.р-16
	У-ПК-10.2	Э, к.р-8, к.р-16
	В-ПК-10.2	Э, к.р-8, к.р-16
ПК-15	З-ПК-15	Э, к.р-8, к.р-16
	У-ПК-15	Э, к.р-8, к.р-16
	В-ПК-15	Э, к.р-8, к.р-16
ПК-4	З-ПК-4	Э, к.р-8, к.р-16
	У-ПК-4	Э, к.р-8, к.р-16
	В-ПК-4	Э, к.р-8, к.р-16
ПК-5	З-ПК-5	Э, к.р-8, к.р-16
	У-ПК-5	Э, к.р-8, к.р-16
	В-ПК-5	Э, к.р-8, к.р-16
ПК-9.2	З-ПК-9.2	Э, к.р-8, к.р-16
	У-ПК-9.2	Э, к.р-8, к.р-16

	В-ПК-9.2	Э, к.р-8, к.р-16
ПК-9.3	З-ПК-9.3	Э, к.р-8, к.р-16
	У-ПК-9.3	Э, к.р-8, к.р-16
	В-ПК-9.3	Э, к.р-8, к.р-16

Шкалы оценки образовательных достижений

Шкала каждого контрольного мероприятия лежит в пределах от 0 до установленного максимального балла включительно. Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет собой сумму баллов, заработанных студентом при выполнении заданий в рамках текущего и промежуточного контроля.

Итоговая оценка выставляется в соответствии со следующей шкалой:

Сумма баллов	Оценка по 4-ех балльной шкале	Оценка ECTS	Требования к уровню освоению учебной дисциплины
90-100	5 – <i>«отлично»</i>	A	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы.
85-89	4 – <i>«хорошо»</i>	B	Оценка «хорошо» выставляется студенту, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос.
75-84		C	
70-74		D	
65-69	3 – <i>«удовлетворительно»</i>	E	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала.
60-64			
Ниже 60	2 – <i>«неудовлетворительно»</i>	F	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

Оценочные средства приведены в Приложении.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ Ф 70 Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : , Москва: ДМК Пресс, 2015
2. ЭИ Ч-45 Применение искусственных нейронных сетей и системы остаточных классов в криптографии : учебное пособие, Москва: Физматлит, 2012
3. ЭИ К 75 Самоорганизующиеся карты : учебное пособие, Москва: Лаборатория знаний, 2017
4. ЭИ М71 Лабораторный практикум по курсу "Введение в теорию нейронных сетей" : , О. А. Мишулина, А. Г. Трофимов, М. В. Щербинина, Москва: МИФИ, 2007

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

1. MATLAB & Simulink Student Version ()

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

1. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. Third Edition. Elsevier, 2012. (http://www.cse.hcmut.edu.vn/~chauvtn/data_mining/Texts/%5B1%5D%20Data%20Mining%20%20Concepts%20and%20Techniques%20%283rd%20Ed%29.pdf)
2. Yoshua Bengio. Learning Deep Architectures for AI // Foundations and Trends in Machine Learning: Vol (<http://www.nowpublishers.com/article/Details/MAL-006>)
3. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, (<http://www.springer.com/us/book/9780387848570>)
4. К. В. Воронцов. Математические методы обучения по прецедентам (теория обучения машин). (<http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>)
5. Bishop C.M. Neural Networks for Pattern Recognition. Oxford, 1994. (http://cs.du.edu/~mitchell/mario_books/Neural_Networks_for_Pattern_Recognition__Christopher_Bishop.pdf)
6. Хайкин С. Нейронные сети: Полный курс. – М.: Издательский дом «Вильямс», 2006. (<http://www.twirpx.com/file/57977/>)
7. Aggarwal, C.C. Data Mining. The Textbook. Springer, XXIX, 2015. (<http://www.springer.com/978-3-319-14141-1>)

<https://online.mephi.ru/>

<http://library.mephi.ru/>

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

9. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

В качестве оценочного средства используется 100 балльная семестровая система, учитывающая, кроме уровня подготовки студента и результатов выполнения курсовой работы, посещаемость лекций и лабораторных занятий, активность (ответы на вопросы при проведении контрольного опроса).

Автор(ы):

Трофимов Александр Геннадьевич, к.т.н.