



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

«СОГЛАСОВАНО»

Руководитель ОП

 Артемьева И.Л.

«УТВЕРЖДАЮ»

И.о. директора департамента

 Смагин С.В.

« 27 _ » сентября 2021 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Машинное обучение и интеллектуальный анализ данных

Направление подготовки 01.04.02 «Прикладная математика и информатика»

(Математическое и программное обеспечение вычислительных машин)

Форма подготовки очная

курс 1,2 семестр 2,3

лекции 36 час.

практические занятия 0 час.

лабораторные работы 36 час.

в том числе с использованием МАО лек. 0 / пр. 0 / лаб. 36 час.

всего часов аудиторной нагрузки – 72 час.

в том числе с использованием МАО – 39 час.

самостоятельная работа 144 час

в том числе на подготовку к экзамену 36 час.

контрольные работы (количество) не предусмотрены

курсовая работа / курсовой проект не предусмотрен

зачет 2 семестр

экзамен 3 семестр

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 01.04.02 Прикладная математика и информатика утвержденного приказом Министерства образования и науки РФ от 10 января 2018 г. №13

Рабочая программа обсуждена на заседании департамента программной инженерии и искусственного интеллекта, протокол № 9.2 от «27» сентября 2021 г.

И.о. директора департамента программной инженерии и искусственного интеллекта к.т.н. Смагин С.В.

Составитель (ли): доцент департамента программной инженерии и искусственного интеллекта Смагин С.В., к.т.н.

Владивосток
2021

Оборотная сторона титульного листа РПД

- I. **Рабочая программа пересмотрена на заседании департамента:**
Протокол от « ____ » _____ 20__ г. № ____
Директор департамента _____
(подпись) (И.О.Фамилия)
- II. **Рабочая программа пересмотрена на заседании департамента:**
Протокол от « ____ » _____ 20__ г. № ____
Директор департамента _____
(подпись) (И.О. Фамилия)

1. Цели и задачи освоения дисциплины:

Цель: обзор основных задач обучения по прецедентам, изучение методов машинного обучения для решения этих задач, а также алгоритмов и программных сред, реализующих эти методы

Задачи:

1. Изучить основные понятия и математические основы машинного обучения, а также примеры прикладных задач обучения по прецедентам в системах искусственного интеллекта.

2. Изучить критерии выбора моделей и методы отбора признаков при решении прикладных задач в системах искусственного интеллекта.

3. Изучить методы классификации (метрические, логические, линейные, байесовские), кластеризации, частичного обучения, регрессии, а также нейронные сети, используемые для решения задач в среде Google Colaboratory.

Для успешного изучения дисциплины у обучающихся должны быть сформированы предварительные компетенции:

- способность к самоорганизации и самообразованию;
- владение основными концепциями, принципами, теориями и фактами, связанными с информатикой;
- способностью осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, представлять ее в требуемом формате с использованием информационных, компьютерных и сетевых технологий;
- владение навыками использования операционных систем, сетевых технологий, средств разработки программного интерфейса, применения языков и методов формальных спецификаций, систем управления базами данных;
- владение навыками использования различных технологий разработки программного обеспечения.

Профессиональные компетенции выпускников и индикаторы их достижения:

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
научно-исследовательский	ПК-1 Способен проводить научные исследования и получать новые научные и прикладные результаты	ПК-1.1 Демонстрирует знание основных достижений и концепций в области прикладной математики и информатики
		ПК-1.2 Использует методы проведения научных

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
	самостоятельно и в составе научного коллектива	исследований и получения новых научных и прикладных результатов самостоятельно и в составе научного коллектива
		ПК-1.3 Самостоятельно и в составе научного коллектива проводит научные исследования
производственно-технологический	ПК-4 Способен разрабатывать и применять математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности	ПК-4.1 Демонстрирует знание языков программирования, алгоритмов, библиотек и пакетов программ, продуктов системного и прикладного программного обеспечения, направлений развития и использования математических и информационных инструментальных средств, автоматизированных систем в научной и практической деятельности
		ПК-4.2 Использует математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности
		ПК-4.3 Применяет методы разработки архитектуры, алгоритмических и программных решений, языки программирования, алгоритмы, библиотеки и пакеты программ в области системного и прикладного программного обеспечения для решения задач научной и проектно-технологической деятельности

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
ПК-1.1 Демонстрирует знание основных достижений и концепций в области прикладной математики и информатики	Знает новые научные результаты и предысторию их появления
	Умеет систематизировать научные результаты, выделять из них главное, и удалять второстепенное
	Владеет навыками сбора и математическими источниками информации
ПК-1.2 Использует методы проведения научных исследований и получения новых научных и прикладных результатов самостоятельно и в составе научного коллектива	Знает классические методы, применяемые в прикладной математике и информатике, необходимые и достаточные условия их реализации
	Умеет самостоятельно выбирать эффективные методы решения поставленных задач и разрабатывать новые методы для получения новых научных и прикладных результатов
	Владеет наукоемкими технологиями и пакетами прикладных программ для решения прикладных задач
ПК-1.3 Самостоятельно и в составе научного коллектива проводит научные исследования	Знает основы стратегического планирования работы коллектива для достижения поставленной цели
	Умеет планировать командную работу, распределять поручения и делегировать полномочия членам команды
	Владеет навыками преодоления возникающих в коллективе разногласий, споров и конфликтов на основе учета интересов всех сторон
ПК-4.1 Демонстрирует знание языков программирования, алгоритмов, библиотек и пакетов программ, продуктов системного и прикладного	Знает концептуальные и теоретические модели классических проблем и задач
	Умеет проводить анализ и обосновывать необходимость работы над данным проектом и оценивать его эффективность

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения)
программного обеспечения, направлений развития и использования математических и информационных инструментальных средств, автоматизированных систем в научной и практической деятельности	Владеет навыками работы над проектами по выбранной тематике
ПК-4.2 Использует математические методы, системное и прикладное программное обеспечение для решения задач научной и проектно-технологической деятельности	Знает современные тенденции и направления в научных исследованиях, проводимых в мире
	Умеет обосновывать и защищать предлагаемый проект, доказывать его эффективность и востребованность на рынке
	Владеет методами построения, анализа и применения математических моделей для оценки состояния, и прогноза развития экономических процессов и явлений; опытом выражения своих мыслей и мнения
ПК-4.3 Применяет методы разработки архитектуры, алгоритмических и программных решений, языки программирования, алгоритмы, библиотеки и пакеты программ в области системного и прикладного программного обеспечения для решения задач научной и проектно-технологической деятельности	Знает современные методы цифровой обработки изображений и средства компьютерной графики
	Умеет выбирать оптимальные системы программирования, наиболее подходящие для решения поставленной задачи
	Владеет навыками работы над производственным проектом в составе группы научных специалистов

2.Трудоёмкость дисциплины и видов учебных занятий по дисциплине

Общая трудоёмкость дисциплины составляет 6 зачётных единиц (216 академических часов).

(1 зачетная единица соответствует 36 академическим часам)

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Лек	Лекции
Лаб	Лабораторные работы
СР	Самостоятельная работа обучающегося в период теоретического обучения
Контроль	Самостоятельная работа обучающегося и контактная работа обучающегося с преподавателем в период промежуточной аттестации

Структура дисциплины:

Форма обучения – очная

Наименование раздела дисциплины	емест	Количество часов по видам учебных занятий и работы обучающегося	Формы промежуточной аттестации

			Лек	Лаб	Пр	ОК	СР	Контроль	
1	Введение, цели и задачи курса, основные термины	2	2		30				УО-1, УО-4, ПР-6 зачет
2	Современная концепция анализа данных, разбор основных типов прикладных задач	2	2						
3	Формальные определения терминов, введение в проблему переобучение	2	2						
4	Метрические методы классификации	2	2	6			36		
5	Логические методы классификации	2	2	4					
6	Линейные методы классификации. Метод стохастического градиента	2	2						
7	Линейные методы классификации. Метод опорных векторов	2	2	4					
8	Байесовские методы классификации	2	2	4					
9	Методы кластеризации и частичного обучения	2	2	4					
10	Метрические методы регрессии	3	2	4					УО-1, УО-4, ПР-6 экзамен
11	Многомерная линейная и нелинейная регрессия	3	2						
12	Критерии выбора моделей и методы отбора признаков	3	2				72	36	
13	Поиск ассоциативных правил	3	2	4					
14	Линейные ансамбли	3	2						
15	Нейронные сети глубокого обучения	3	8	6					
	Итого:		36	36			108	36	Зачет, экзамен

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Трудоемкость теоретической части курса 36 час.

1 семестр (18 час.)

Тема 1. Введение, цели и задачи курса, основные термины (2 час.)

Рассматриваются понятия: модель предметной области, задача обучения по прецедентам, алгоритм обучения, решающее правило, критерий качества обучения, обучающая и контрольная выборки, модельные и реальные данные,

объекты, признаки, свойства, классы, кластеры, типы шкал (бинарные, номинальные, порядковые, количественные).

Тема 2. Современная концепция анализа данных, разбор основных типов прикладных задач (2 час.)

Рассматриваются типы задач обучения с учителем: классификация, регрессия, ранжирование, прогнозирование. Рассматриваются типы задач обучения без учителя: кластеризация, поиск ассоциативных правил, фильтрация выбросов, сокращение размерности, заполнение пропущенных значений. Каждый тип задач сопровождается одним или несколькими детально расписанными примерами.

Тема 3. Формальные определения терминов, введение в проблему переобучения (2 час.)

Рассматриваются формальные (математические) определения терминов: модель алгоритмов, алгоритм обучения, этапы обучения, ошибка, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль, проблема переобучения.

Тема 4. Метрические методы классификации (2 час.)

Рассматриваются понятия: гипотеза компактности, метрика, метод k ближайших соседей KNN и его обобщения, окно Парзена ParzenWindows и потенциальные функции, отбор эталонов (понятие отступа, алгоритм отбора эталонных объектов STOLP), отбор признаков и оптимизация метрики (задача выбора метрики, жадный алгоритм отбора признаков, полный скользящий контроль CCV).

Тема 5. Логические методы классификации (2 час.)

Рассматриваются понятия: закономерность, информативность, интерпретируемость, решающие деревья (алгоритм ID3, небрежные решающие деревья ODT, бинаризация данных).

Тема 6. Линейные методы классификации. Метод стохастического градиента (2 час.)

Рассматриваются понятия: минимизация эмпирического риска для градиентных методов обучения, линейный классификатор, метод стохастического градиента SG.

Тема 7. Линейные методы классификации. Метод опорных векторов (2 час.)

Рассматриваются понятия: принцип оптимальной разделяющей гиперплоскости, двойственная задача, понятие опорного вектора, обобщения

линейного SVM (ядра и спрямляющие пространства, нейронные сети и SVM, обзор регуляризаторов для SVM).

Тема 8. Байесовские методы классификации (2 час.)

Рассматриваются понятия: оптимальный байесовский классификатор, вероятностная постановка задачи классификации, задача восстановления плотности распределения, наивный байесовский классификатор NaiveBayes.

Тема 9. Методы кластеризации и частичного обучения (2 час.)

Рассматриваются понятия: цели кластеризации и частичного обучения, некорректность задачи кластеризации, типы кластерных структур, проблема чувствительности к выбору метрики, графовые методы кластеризации (алгоритм выделения связанных компонент, алгоритм ФОРЭЛ, функционалы качества кластеризации), иерархическая кластеризация (таксономия) (агломеративная иерархическая кластеризация, дендрограмма и свойство монотонности, свойства сжатия, растяжения и редуктивности).

2 семестр (18 час.)

Тема 10. Метрические методы регрессии (2 час.)

Рассматриваются понятия: непараметрическая регрессия.

Тема 11. Многомерная линейная и нелинейная регрессия (2 час.)

Рассматриваются понятия: задача регрессии, многомерная линейная регрессия, метод наименьших квадратов, его вероятностный смысл и геометрический смысл, проблемы мультиколлинеарности и переобучения.

Тема 12. Критерии выбора моделей и методы отбора признаков (2 час.)

Рассматриваются понятия: критерии качества классификации (чувствительность и специфичность, ROC-кривая и AUC, точность и полнота), внутренние и внешние критерии, эмпирические и аналитические критерии.

Тема 13. Поиск ассоциативных правил (2 час.)

Рассматриваются понятия: ассоциативное правило и его связь с понятием логической закономерности, примеры прикладных задач (анализ рыночных корзин, выделение терминов и тематики текстов), алгоритм APriori.

Тема 14. Линейные ансамбли (2 час.)

Рассматриваются понятия: базовый алгоритм, корректирующая операция, простое голосование (комитет большинства), стохастические методы (бэггинг и метод случайных подпространств), случайный лес (Random Forest), взвешенное голосование, преобразование простого голосования во взвешенное.

Тема 15. Нейронные сети глубокого обучения (8 час.)

Рассматриваются понятия: сверточные нейронные сети (CNN) для изображений, сверточный нейрон, pooling нейрон, выборка размеченных изображений ImageNet, сверточные сети (для сигналов, текстов, графов, игр),

рекуррентные нейронные сети (RNN), обучение рекуррентных сетей (Backpropagation Through Time (BPTT)), сети долгой кратковременной памяти (Long short-term memory, LSTM), рекуррентная сеть Gated Recurrent Unit (GRU), векторные представления дискретных данных, перенос обучения (transfer learning), самообучение (self-supervised learning).

II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА И САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Практические занятия (0 час.)

Лабораторные работы (36 час.)

1 семестр (18 час.)

Лабораторная работа №1. Метрические методы классификации (6 час.)

Настройка значений параметров методов KNN и ParzenWindows для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №2. Логические методы классификации (4 час.)

Настройка значений параметров методов ID3 и ODT для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №3. Линейные методы классификации (4 час.)

Настройка значений параметров методов SG и SVM для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №4. Байесовские методы классификации (4 час.)

Настройка значений параметров метода NaiveBayes для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

2 семестр (18 час.)

Лабораторная работа №5. Методы кластеризации (4 час.)

Настройка значений параметров для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №6. Методы регрессии (4 час.)

Настройка значений параметров для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №7. Поиск ассоциативных правил (4 час.)

Настройка значений параметров для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Лабораторная работа №8. Нейронные сети глубокого обучения (6 час.)

Настройка значений параметров для исследования заданной обучающей выборки с UCI ML Repository в среде Google Colaboratory.

Ш. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине включает в себя:

- план-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию;
- требования к представлению и оформлению результатов самостоятельной работы;
- критерии оценки выполнения самостоятельной работы.

План-график выполнения самостоятельной работы по дисциплине

/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
	4 неделя обучения	Изучение основных понятий машинного обучения	6	Собеседование
	6 неделя обучения	Изучение математических основ машинного обучения	6	Собеседование
	8 неделя обучения	Рассмотрение примеров прикладных задач обучения по прецедентам в системах искусственного интеллекта	6	Собеседование
	10 неделя обучения	Изучение критериев выбора моделей и методов отбора признаков при решении прикладных задач в системах искусственного интеллекта	6	Собеседование
	14 неделя обучения	Выбор обучающей выборки в UCI Machine Learning Repository	6	Собеседование
	18 неделя обучения	Изучение методов классификации и	6	Проверка отчетов, собеседование

		регрессии		
	22 неделя обучения	Изучение методов кластеризации и частичного обучения	12	Проверка отчетов, собеседование
	26 неделя обучения	Изучение нейронных сетей	12	Проверка отчетов, собеседование
	32 неделя обучения	Исследование заданными методами машинного обучения выборки в среде Google Colaboratory	48	Проверка отчетов, собеседование
	34 -36 недели обучения	Подготовка к экзамену	36	Проверка отчетов, собеседование
Итого:			144	

Рекомендации по самостоятельной работе студентов

Самостоятельная работа обучающихся подразумевает обязательную подготовку к лабораторным занятиям (оформление отчетов), изучение основной и дополнительно литературы по дисциплине, подготовку к текущему контролю и промежуточной аттестации в конце семестра, консультации преподавателей

Рекомендации по работе с литературой

Для более эффективного освоения и усвоения материала рекомендуется ознакомиться с теоретическим материалом по той или иной теме до проведения лабораторного занятия. Всю учебную литературу желательно изучать «под конспект».

Цель написания конспекта по дисциплине – сформировать навыки по поиску, отбору, анализу и формулированию учебного материала.

Работу с теоретическим материалом по теме можно проводить по следующей схеме:

- название темы;
- цели и задачи изучения темы;
- основные вопросы темы;
- характеристика основных понятий и определений, необходимых для усвоения данной темы;
- краткие выводы, ориентирующие на определенную совокупность сведений, основных идей, ключевых положений, систему доказательств, которые необходимо усвоить.

При работе над конспектом обязательно выявляются и отмечаются трудные для самостоятельного изучения вопросы, с которыми уместно обратиться к преподавателю при посещении консультаций, либо в индивидуальном порядке.

Подготовка к лабораторным работам

Подготовку к каждой лабораторной работе или к практическому занятию каждый

студент должен начать с изучения теоретического материала и ознакомления с планом, который отражает содержание предложенной темы. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы по теме задания, правильном его выполнении.

В процессе выполнения лабораторной работы или практического задания студент должен создать требуемый документ с помощью предлагаемого программного средства и выполнить требуемые в задании операции. Задание по лабораторной или практической работе содержит методические указания по подготовке документа, который должен быть получен в результате выполнения работы. При подготовке следует их внимательно прочесть.

Критерии оценки лабораторных работ

– 100-86 - выполнены все задания лабораторной работы, студент четко и без ошибок ответил на все контрольные вопросы.

– 85-76 - выполнены все задания лабораторной работы; студент ответил на все контрольные вопросы с замечаниями.

– 75-61 выполнены все задания лабораторной работы с замечаниями; студент ответил на все контрольные вопросы с замечаниями.

- 60-50 баллов - студент не выполнил или выполнил неправильно задания лабораторной работы; студент ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы.

Подготовка презентации и доклада

Для подготовки презентации рекомендуется использовать: PowerPoint, MS Word, Acrobat Reader, LaTeX-овский пакет beamer. Последовательность подготовки презентации:

1. Четко сформулировать цель презентации: вы хотите свою аудиторию мотивировать, убедить, заразить какой-то идеей или просто формально отчитаться.

2. Определить каков будет формат презентации: живое выступление (тогда, сколько будет его продолжительность) или электронная рассылка (каков будет контекст презентации).

3. Отобрать всю содержательную часть для презентации и выстроить логическую цепочку представления.

4. Определить ключевые моменты в содержании текста и выделить их.

5. Определить виды визуализации (иллюстрации, образы, диаграммы, таблицы) для отображения их на слайдах в соответствии с логикой, целью и спецификой материала.

6. Подобрать дизайн и форматировать слайды (количество картинок и текста, их расположение, цвет и размер).

7. Проверить визуальное восприятие презентации.

Практические советы по подготовке презентации - готовьте отдельно:

- печатный текст + слайды + раздаточный материал;
- *слайды* – визуальная подача информации, которая должна содержать минимум текста, максимум изображений, несущих смысловую нагрузку, выглядеть наглядно и просто;
- *текстовое содержание презентации* – устная речь или чтение, которая должна включать аргументы, факты, доказательства и эмоции;
- *рекомендуемое число слайдов* 17-22;
- *обязательная информация для презентации*: тема, фамилия и инициалы выступающего; план сообщения; краткие выводы из всего сказанного; список использованных источников;
- *раздаточный материал* – должен обеспечивать ту же глубину и охват, что и живое выступление: люди больше доверяют тому, что они могут унести с собой, чем исчезающим изображениям, слова и слайды забываются, а раздаточный материал остается постоянным осязаемым напоминанием; раздаточный материал важно раздавать в конце презентации; раздаточный материалы должны отличаться от слайдов, должны быть более информативными.

	50-60 баллов (неудовлетворительно)	61-75 баллов (удовлетворительно)	76-85 баллов (хорошо)	86-100 баллов (отлично)
критерии	Содержание критериев			
Раскрытие проблемы	Проблема не раскрыта. Отсутствуют выводы	Проблема раскрыта не полностью. Выводы не сделаны и/или выводы не обоснованы	Проблема раскрыта. Проведен анализ проблемы без привлечения дополнительной литературы. Не все выводы сделаны и/или обоснованы	Проблема раскрыта полностью. Проведен анализ проблемы с привлечением дополнительной литературы Выводы обоснованы
Представление	Представляемая информация логически не связана. профессиональные термины Студент демонстрирует неумение использовать понятийный аппарат	Представляемая информация не систематизирована и/или не последовательна. Студент демонстрирует затруднения с использованием научно-понятийного аппарата и терминологии учебной дисциплины;	Представляемая информация не систематизирована и последовательна. Студент демонстрирует умение пользоваться концептуально-понятийным аппаратом в процессе анализа основных проблем в рамках данной темы	Представляемая информация систематизирована, последовательна и логически связана. Студент демонстрирует отчетливое и свободное владение концептуально-понятийным аппаратом, научным языком и терминологией соответствующей научной области.
Ответы на вопросы	Нет ответов на вопросы	Только ответы на элементарные вопросы	Ответы на вопросы полные и/или частично полные	Ответы на вопросы полные, с приведением примеров и/или пояснений
Оформление	Не использованы технологии Power Point. Больше 4 ошибок в представляемой информации	Использованы технологии Power Point частично. 3-4 ошибки в представляемой информации	Использованы технологии Power Point. Не более 2 ошибок в представляемой информации	Широко использованы технологии (Power Point и др.).Отсутствуют ошибки в представляемой информации

Критерии оценки презентации доклада

Менее 60 баллов	незачтено	неудовлетворительно
От 61 до 75 баллов	зачтено	удовлетворительно
От 76 до 85 баллов	зачтено	хорошо
От 86 до 100 баллов	зачтено	отлично

IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы/темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства - наименование		
			текущий контроль	промежуточная аттестация	
1.	Темы 1-3	ПК-1	Знает	Собеседование УО1, круглый стол УО-4	Экзамен Вопросы 1-12 (теоретическая часть)
			Умеет Владеет	Л/работы 1-8, ПР-6	
2.	Темы 4-9	ПК-1, ПК-4	Знает	Собеседование УО1, круглый стол УО-4	Экзамен Вопросы 1-13 (практическая часть)
			Умеет Владеет	Л/работы 1-4, ПР-6	
3.	Темы 10-15	ПК-1 ПК-4	Знает	Собеседование УО1, круглый стол УО-4	Экзамен Вопросы 14-20 (практическая часть)
			Умеет Владеет	Л/работы 5-8, ПР-6	

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, а также качественные критерии оценивания, которые описывают уровень сформированности компетенций, представлены в разделе VIII.

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

1. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных – М.: ДМК Пресс,

2015. – 400 с. <http://www.studentlibrary.ru/book/ISBN9785970602737.html>
<https://e.lanbook.com/book/69955>
2. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2-е издание / пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 302 с.: ил.
<http://www.studentlibrary.ru/book/ISBN9785970603307.html>
 3. Вьюгин, В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс]: учеб. пособие – М.: МЦНМО, 2013. – 304 с. <https://e.lanbook.com/book/56397>
 4. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс]: рук. / С. Рашка; пер. с англ. Логунова А.В. – М.: ДМК Пресс, 2017. – 418 с. <https://e.lanbook.com/book/100905>
 5. Кулаичев А.П. Методы и средства комплексного анализа данных: учебное пособие для вузов. – М.: [Форум]: ИНФРА-М, 2014. 511 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:795113&theme=FEFU>

Дополнительная литература
(печатные и электронные издания)

1. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999. – 270 с.
<https://lib.dvfu.ru:8443/lib/item?id=chamo:10172&theme=FEFU>
2. Кухаренко Б.Г. Интеллектуальные системы и технологии – М.: Московская государственная академия водного транспорта, 2015. – 116 с. <http://www.iprbookshop.ru/47933.html>
3. Давнис В.В., Тинякова В.И., Мокшина С.И., Алексеева А.И. Компьютерные решения задач многомерной статистики. Часть 1. Кластерный и дискриминантный анализ: Учебное пособие. – Воронеж: Изд-во ВГУ, 2005. – 37 с. <http://window.edu.ru/resource/417/40417>
4. Симчера В.М. Методы многомерного анализа статистических данных: учеб. пособие. – М.: Финансы и статистика, 2008. – 400 с.: ил.
<http://www.studentlibrary.ru/book/ISBN9785279031849.html>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <http://machinelearning.ru/> MachineLearning.ru Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. http://shad.yandex.ru/lectures/machine_learning.xml Видеолекции курса «Машинное обучение» Школы анализа данных.
3. <http://www.studentlibrary.ru/book/ISBN9785940575061.html> Николенко С.И., Тулупьев А.Л. Самообучающиеся системы. – М.: МЦНМО, 2009. – 288 с.: 24 ил.
4. <http://znanium.com/catalog.php?bookinfo=451186> Информационные аналитические системы: учебник / Т.В. Алексеева, Ю.В. Амириди, В.В. Дик и др.; под ред. В.В. Дика. – М.: МФПУ Синергия, 2013. – 384 с. – (Университетская серия). – ISBN 978-5-4257-0092-6.

Перечень информационных технологий и программного обеспечения

Для проведения практических занятий требуется следующее программное обеспечение: Google Colaboratory, CPython.

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Планирование и организация времени, отведенного на изучение дисциплины. Приступить к освоению дисциплины следует незамедлительно в самом начале учебного семестра. Рекомендуется изучить структуру и основные положения Раб очей программы дисциплины. Обратит внимание, что кроме аудиторной работы (лекции, лабораторные занятия) планируется самостоятельная работа, итоги которой влияют на окончательную оценку по итогам освоения учебной дисциплины. Все задания (аудиторные и самостоятельные) необходимо выполнять и предоставлять на оценку в соответствии с графиком.

В процессе изучения материалов учебного курса предлагаются следующие формы работ: чтение лекций, лабораторные занятия, задания для самостоятельной работы.

Лекционные занятия ориентированы на освещение вводных тем в каждый раздел курса и призваны ориентировать студентов в предлагаемом

материале, заложить научные и методологические основы для дальнейшей самостоятельной работы студентов.

Лабораторные занятия акцентированы на наиболее принципиальных и проблемных вопросах курса и призваны стимулировать выработку практических умений.

Особо значимой для профессиональной подготовки студентов является *самостоятельная работа* по курсу. В ходе этой работы студенты отбирают необходимый материал по изучаемому вопросу

Освоение курса способствует развитию навыков обоснованных и самостоятельных оценок фактов и концепций. Поэтому во всех формах контроля знаний, особенно при сдаче зачета, внимание обращается на понимание проблематики курса, на умение практически применять знания и делать выводы.

Работа с литературой. Рекомендуется использовать различные возможности работы с литературой: фонды научной библиотеки ДВФУ и электронные библиотеки (<http://www.dvfu.ru/library/>), а также доступные для использования другие научно-библиотечные системы.

Подготовка к экзамену/зачету. К сдаче экзамена/зачета допускаются обучающиеся, выполнившие все задания (лабораторные, самостоятельные), предусмотренные учебной программой дисциплины, посетившие не менее 85% аудиторных занятий.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

Материально-техническое и программное обеспечение дисциплины

Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
690922, Приморский край, г. Владивосток, остров Русский, полуостров Саперный,	Помещение укомплектовано специализированной учебной мебелью (посадочных мест – 13) Оборудование:	1С Предприятие8 (8.2), 7-Zip, ABBYY Lingvo12,Alice 3, Anaconda3,Autodesk,CodeBlocks,CorelDRAW X7,Dia,Directum4.8,DosBox-0.74,Farmanager,Firebird 2.5,FlameRobin,Foxit Reader,Free Pascal,Geany,Ghostscript,Git,Greenfoot,gsview,Inscapе0.91,Java,Java development Kit,Kaspersky,Lazarus,LibreOffice4.4,MatLab R2017b,Maxima 5.37.2,Microsoft Expression,Microsoft Office 2013,Microsoft Silverlight,Microsoft Silverlight 5SDK-русский,MicrosoftSistem Center,Microsoft Visial Studio

<p>поселок Аякс, 10, корпус D, ауд. D 733,733а. Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации</p>	<p>ЖК-панель 47", Full HD, LG M4716 CCBA – 1 шт. Доска аудиторная, Моноблок Lenovo C360G-i34164G500UDK с лицензионными программами Microsoft Office 2013(13 шт.) и аудиовизуальными средствами проектор Panasonic DLPProjectorPT-D2110XE</p>	<p>2012,MikTeX2.9,MySQL,NetBeans,Notepad++,Oracle VM VirtualBox,PascalABC.NET,PostgreSQL 9.4,PTC Mathcad,Putty,PyQt GPL v5.4.1 for Pythonv 3.4,Python2.7(3.4,3.6),QGIS Brighton,RStudio,SAM CoDeC Pack,SharePoint,Strawberry Perl,Tecnomatix,TeXnicCenter,TortoiseSVN,Unity2017.3.1f1,Veusz,Vim8.1,Visual Paradigm CE,Visual Studio2013,Windows Kits,Windows Phone SDK8.1,Xilinx Design ToolsAcrobat ReaderDC,AdobeBridge CS3,AdobeDeviceCentralCS3,Adobe ExtendScript Toolkit 2,Adobe Photoshpe CS3,DVD-студия Windows,GoogleChrome,Internet Explorer,ITMOproctor,Mozilla Firefox, Visual Studio Installer,Windows Media Center, WinSCP,</p>
--	--	---

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.

VIII. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

Используются следующие оценочные средства:

Устный опрос:

1. Собеседование (УО-1)
2. Презентация / сообщение (УО-3)
3. Круглый стол (УО-4)

Письменные работы:

1. Лабораторная работа (ПР-6)
2. Тест (ПР-1)

Устный опрос

Устный опрос позволяет оценить знания и кругозор студента, умение логически построить ответ, владение монологической речью и иные коммуникативные навыки.

Обучающая функция состоит в выявлении деталей, которые по каким-то причинам оказались недостаточно осмысленными в ходе учебных занятий и при подготовке к зачёту.

Собеседование (УО-1) – средство контроля, организованное как специальная беседа преподавателя с обучающимся на темы, связанные с

изучаемой дисциплиной, и рассчитанное на выяснение объема знаний обучающегося по определенному разделу, теме, проблеме и т.п.

Презентация / сообщение (УО-3) – продукт самостоятельной работы обучающегося, представляющий собой публичное выступление по представлению полученных результатов решения определенной учебно-практической, учебно-исследовательской или научной темы.

Круглый стол (УО-4) - Оценочные средства, позволяющие включить обучающихся в процесс обсуждения спорного вопроса, проблемы и оценить их умение аргументировать собственную точку зрения.

Письменные работы

Письменная проверка используется во всех видах контроля и осуществляется как в аудиторной, так и во внеаудиторной работе.

Лабораторная работа (ПР-6) – средство для закрепления и практического освоения материала по определенному разделу.

Тест (ПР-1) - Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.

Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины

Оценочные средства для промежуточной аттестации

Промежуточная аттестация студентов по дисциплине проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной. Форма отчётности по дисциплине – экзамен (3-й семестр), зачет (4-й, весенний семестр).

Экзамен проводится в письменной форме с использованием защиты проекта.

Критерии выставления оценки студенту

Баллы (рейтинговой оценки)	Оценка зачета/ экзамена (стандартная)	Требования к сформированным компетенциям
---	--	---

86-100	«зачтено»/ «отлично»	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, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76-85	«зачтено»/ «хорошо»	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.
61-75	«зачтено»/ «удовлетворительно»	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, испытывает затруднения при выполнении практических работ.
0-60	«не зачтено»/ «неудовлетворительно»	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

Вопросы к экзамену

Теоретическая часть:

1. Определение машинного обучения как направления исследований, его цели и задачи, особенности, основные области применения методов.
2. Типы обучения (дедуктивное, индуктивное, комбинированное).
Типичный сценарий индуктивного обучения.

3. Признаковое описание объектов обучающей и контрольной выборок. Основные типы признаков и критерии их отбора.
4. Задача обучения по прецедентам, обобщающая способность, метод и алгоритм обучения, решающее правило, этапы обучения и применения.
5. Критерий качества обучения (функционал качества), ошибка, эмпирический риск, сведение задачи обучения к задаче оптимизации.
6. Классификатор, ошибки 1-ого и 2-ого рода, проблема переобучения, задача чемпионата мира среди алгоритмов классификации.
7. Основные типы задач обучения с учителем (классификация, регрессия, ранжирование, прогнозирование), их постановки и отличия.
8. Основные типы задач обучения без учителя (кластеризация, поиск ассоциативных правил, фильтрация выбросов, сокращение размерности, заполнение пропущенных значений), их постановки и отличия.
9. Другие типы задач: частичное обучение, обучение с подкреплением, динамическое обучение, активное обучение, метаобучение.
10. Примеры задачи классификации: медицинская диагностика, оценивание заемщиков, предсказание ухода клиента. Пример задачи восстановления регрессии: прогнозирование потребительского спроса.
11. Пример задачи ранжирования: ранжирование текстовых документов. Примеры задачи кластеризации: рубрикация текстов, рекомендации.
12. Оценка качества классификации: TP, TN, FP, FN; ROC-кривая, AUC-ROC. Точность и полнота для двухклассовой и многоклассовой классификации.

Практическая часть:

1. Метрические методы классификации: гипотеза компактности, метрика, метрическое пространство, обобщенный метрический классификатор.
2. Метрические методы классификации: метод ближайшего соседа, метод k ближайших соседей, метод окна Парзена, метод потенциальных функций.
3. Метрические методы классификации: понятие отступа, типы объектов (в зависимости от отступа), алгоритм STOLP.
4. Логические методы классификации: логическая закономерность, интерпретируемость, информативность, часто используемые виды закономерностей, часто используемые критерии информативности, обобщенный алгоритм нахождения информативных закономерностей.
5. Логические методы классификации: определение бинарного решающего дерева, пример решающего дерева, жадный алгоритм построения дерева ID3, достоинства и недостатки решающих деревьев ID3.

6. Логические методы классификации: определение бинарного решающего дерева, редукция дерева, небрежные решающие деревья ODT, задача бинаризации вещественного признака.
7. Линейные методы классификации (метод стохастического градиента): задача построения разделяющей поверхности, линейный классификатор, математическая модель нейрона, градиентный метод численной минимизации, алгоритм SG.
8. Линейные методы классификации (метод опорных векторов): кусочно-линейная аппроксимация, задача SVM, оптимальная разделяющая гиперплоскость, достоинства и недостатки метода.
9. Байесовские методы классификации: байесовский классификатор апостериорная вероятность, вероятностная постановка задачи, функционал среднего риска, оптимальный байесовский классификатор.
10. Байесовские методы классификации: разбиение задачи классификации на две подзадачи и алгоритм решение первой из них, три основных подхода к оцениванию плотностей, наивный байесовский классификатор.
11. Методы кластеризации: постановка задачи кластеризации, некорректность задачи кластеризации, цели кластеризации, типы кластерных структур, проблема чувствительности к выбору метрики.
12. Методы кластеризации: графовые методы кластеризации (алгоритм выделения связных компонент, алгоритм кратчайшего незамкнутого пути, алгоритм ФОРЭЛ, функционалы качества кластеризации).
13. Методы кластеризации: агломеративная иерархическая кластеризация (алгоритмы Ланса-Уильямса (обычный и редуцированный), формула Ланса-Уильямса и ее частные случаи, визуализация кластерной структуры).
14. Метрические методы регрессии: непараметрическая регрессия.
15. Многомерная линейная и нелинейная регрессия: задача регрессии, метод наименьших квадратов, его вероятностный смысл и геометрический смысл, проблемы мультиколлинеарности и переобучения.
16. Критерии выбора моделей и методы отбора признаков: критерии качества классификации (чувствительность и специфичность, ROC-кривая и AUC, точность и полнота), внутренние и внешние критерии, эмпирические и аналитические критерии.
17. Поиск ассоциативных правил: ассоциативное правило и его связь с понятием логической закономерности, примеры прикладных задач, алгоритм APriori.
18. Линейные ансамбли: базовый алгоритм, корректирующая операция, простое голосование (комитет большинства), стохастические методы,

случайный лес (Random Forest), взвешенное голосование, преобразование простого голосования во взвешенное.

19. Нейронные сети глубокого обучения: сверточные нейронные сети (CNN) для изображений, сверточный нейрон, pooling нейрон, выборка размеченных изображений ImageNet, сверточные сети для сигналов, текстов, графов, игр.
20. Нейронные сети глубокого обучения: рекуррентные нейронные сети (RNN), обучение рекуррентных сетей, сети долгой кратковременной памяти, рекуррентная сеть GRU, векторные представления дискретных данных, перенос обучения, самообучение.

Оценочные средства для текущей аттестации

Текущая аттестация студентов по дисциплине проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной.

Текущая аттестация проводится в форме контрольных мероприятий (собеседования, презентации, лабораторных работ, тестов) по оцениванию фактических результатов обучения студентов.

Объектами оценивания выступают:

- степень усвоения теоретических знаний - оценивается в форме собеседования и тестирования;
- уровень овладения практическими умениями и навыками – оценивается в форме защиты лабораторных работ.

Составляется календарный план контрольных мероприятий по дисциплине. Оценка посещаемости, активности обучающихся на занятиях, своевременность выполнения различных видов заданий ведётся на основе журнала, который ведёт преподаватель в течение учебного семестра.

Критерии оценки теста

Оценка	Требования
<i>отлично</i>	Более 80% правильных ответов
<i>хорошо</i>	65 – 80% правильных ответов
<i>удовл</i>	50 - 65% правильных ответов
<i>неудовл</i>	Менее 50% правильных ответов

Критерии оценки практических занятий

- 100-86 баллов выставляется, если студент/группа точно определили содержание и составляющие части задания, умеют аргументированно отвечать на вопросы, связанные с заданием.

Продемонстрировано знание и владение навыками самостоятельной исследовательской работы по теме. Фактических ошибок, связанных с пониманием проблемы, нет.

- 85-76 - баллов - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.
- 75-61 балл – проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы
- 60-50 баллов - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы

Шкала оценивания

Менее 60 баллов	Незачтено	неудовлетворительно
От 61 до 75 баллов	Зачтено	удовлетворительно
От 76 до 85 баллов	Зачтено	хорошо
От 86 до 100 баллов	Зачтено	отлично