

Федеральное агентство связи
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»
(СибГУТИ)
Уральский технический институт связи и информатики (филиал) в г. Екатеринбурге
(УрТИСИ СибГУТИ)



РАБОЧАЯ ПРОГРАММА

по дисциплине «**Методы машинного обучения**»
для основной профессиональной образовательной программы по направлению 09.03.01
«Информатика и вычислительная техника»
направленность (профиль) - Программное обеспечение средств вычислительной техники и
автоматизированных систем
квалификация - бакалавр
форма обучения - очная, заочная
год начала подготовки (по учебному плану) - 2020

Екатеринбург 2020

Федеральное агентство связи
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»
(СибГУТИ)
Уральский технический институт связи и информатики (филиал) в г. Екатеринбурге
(УрТИСИ СибГУТИ)

Утверждаю
Директор УрТИСИ СибГУТИ
_____ Е.А. Минина
« _____ » _____ 2020 г.

РАБОЧАЯ ПРОГРАММА

по дисциплине **«Методы машинного обучения»**
для основной профессиональной образовательной программы по направлению
09.03.01 «Информатика и вычислительная техника»
направленность (профиль) – Программное обеспечение средств вычислительной техники и
автоматизированных систем
квалификация – бакалавр
форма обучения – очная, заочная
год начала подготовки (по учебному плану) – 2020

Екатеринбург 2020

1. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к обязательной части учебного плана. Шифр дисциплины в учебном плане – Б1.В.ДВ.01.01

<i>ОПК–1– Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности</i>	
Предшествующие дисциплины и практики	Информатика, Математика, Алгебра и геометрия, Физика, Математика, Физика, Математическая логика и теория алгоритмов, Дискретная математика, Операционные системы, Теория вероятностей и математическая статистика, Специальные главы математического анализа, Вычислительная математика, Теория массового обслуживания
Дисциплины и практики, изучаемые одновременно с данной дисциплиной	-
Последующие дисциплины и практики	Теория информации
<i>ПК–1–Способен разрабатывать требования и проектировать программное обеспечение</i>	
Предшествующие дисциплины и практики	Программирование, Технологии разработки программного обеспечения, Алгоритмы и вычислительные методы оптимизации, Функциональное и логическое программирование,
Дисциплины и практики, изучаемые одновременно с данной дисциплиной	Современные технологии программирования
Последующие дисциплины и практики	Современные технологии программирования

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате освоения дисциплины обучающийся должен демонстрировать освоение следующих компетенций по дескрипторам «знания, умения, владения», соответствующие тематическим разделам дисциплины, и применимые в их последующем обучении и профессиональной деятельности:

ОПК-1 – *Способен применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности.*

Знать:

– основы вычислительной техники и программирования

Уметь:

– решать стандартные профессиональные задачи с применением общетехнических знаний, моделирования.

Иметь навыки:

– теоретического и экспериментального исследования объектов профессиональной деятельности.

ПК-1 – *Способен разрабатывать требования и проектировать программное обеспечение*

Знать:

– формализацию задачи машинного обучения;

– понятие больших данных и их свойства;

– постановку задачи классификации и регрессии;

– понятие обобщенного метрического классификатора;

– алгоритмы метрической классификации;

– основные принципы построения логических алгоритмов классификации;

– линейные методы классификации.

Уметь:

– использовать алгоритмы обработки информации для различных приложений;

– выполнять постановку задачи машинного обучения;

– выбирать методы и средства для решения задач машинного обучения;

Иметь навыки:

– использования инструментальных средств решения задач машинного обучения;

– применения методов интеллектуального анализа информации.

3 ОБЪЁМ ДИСЦИПЛИНЫ

3.1 Очная форма обучения

Общая трудоемкость дисциплины, изучаемой в 7 семестре, составляет 3 зачетных единиц. По дисциплине предусмотрен *зачет*

Виды учебной работы	Всего часов/зачетных единиц	Семестр
		7
Аудиторная работа (всего)	66/1.83	66/1.83
В том числе в интерактивной форме	2/0.06	2/0.06
Лекции (ЛК)	34/0.94	34/0.94
Лабораторные работы (ЛР)	-	-
Практические занятия (ПЗ)	32/0.88	32/0.88
Самостоятельная работа студентов (всего)	33/0.91	33/0.91
Контроль	9/0.25	9/0.25
Проработка лекций	10/0.28	10/0.28
Подготовка к практическим занятиям и оформление отчетов	10/0.28	10/0.28
Подготовка к лабораторным занятиям и оформление отчетов	-	-
Выполнение курсового проекта	-	-
Подготовка и сдача зачета	13/0.36	13/0.36
Общая трудоемкость дисциплины, часов	108/3	108/3

Одна зачетная единица (ЗЕ) эквивалентна 36 часам.

3.2 Заочная форма обучения

Общая трудоемкость дисциплины, изучаемой на 5 курсе, составляет 9 зачетных единиц. По дисциплине предусмотрен *расчетно-графическая работа, зачет*

Виды учебной работы	Всего часов/зачетных единиц	Курс
		5
Аудиторная работа (всего)	20/0.56	20/0.56
В том числе в интерактивной форме	12/0.33	12/0.33
Лекции (ЛК)	8/0.22	8/0.22
Лабораторные работы (ЛР)	-	-
Практические занятия (ПЗ)	12/0.33	12/0.33

ПК	-	-
Самостоятельная работа студентов (всего)	84/2.33	84/2.33
Контроль	4/0.11	4/0.11
Проработка лекций	15/0.42	15/0.42
Подготовка к практическим занятиям и оформление отчетов	15/0.42	15/0.42
Подготовка к лабораторным занятиям и оформление отчетов	-	-
Выполнение РГР	30/0.83	30/0.83
Подготовка и сдача зачета	24/0.67	24/0.67
Общая трудоемкость дисциплины, часов	108/3	108/3

Одна зачетная единица (ЗЕ) эквивалентна 36 часам.

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ «СТРУКТУРЫ И АЛГОРИТМЫ ОБРАБОТКИ ДАННЫХ» ПО ВИДАМ УЧЕБНЫХ ЗАНЯТИЙ

4.1 Содержание лекционных занятий

№ раздела дисциплины	Наименование лекционных тем (разделов) дисциплины и их содержание	Объем в часах	
		О	З
1	<p>Тема 1 Введение в машинное обучение</p> <p>Основные понятия. Определение предмета машинного обучения. Примеры задач и областей приложения. Образы и признаки. Типы задач предсказания. Регрессия. Таксономия. Классификация. Типы ошибок классификации. Обобщающая способность классификатора. Принцип минимизации эмпирического риска. Недообучение. Переобучение. Статистический, нейросетевой и структурно-лингвистический подходы к распознаванию образов. Структура типичной системы распознавания образов. Цикл построения системы распознавания образов.</p>	4	2
2	<p>Тема 2 Классификация. Общие принципы. Этапы классификации.</p> <p>Алгоритмы обучения классификаторов с учителем и без учителя. Дискриминантный анализ. Геометрическая интерпретация задачи классификации. Проективный подход. Метрики в пространстве признаков. Евклидово расстояние. Расстояние Махаланобиса. Ошибки первого и второго рода. Чувствительность и избирательность. Кривая мощностикритерия классификации. ROC-кривые. Проверка классификатора. Проверка тестовой выборкой. Перекрестная проверка. Оценка информативности признаков.</p>	10	2
3	<p>Тема 3 Основные методы машинного обучения</p> <p>Байесовская классификация. Условная вероятность. Формула полной вероятности. Формула Байеса. Статистическое распознавание образов. Наивный байесовский классификатор. Задача классификации спама. Критерий отношения правдоподобия. Байесовский уровень ошибки. Байесовский риск. Критерий Байеса. Максимальный апостериорный критерий. Критерий максимального правдоподобия. Многоклассовые байесовские классификаторы. Байесовские классификаторы для нормально распределенных классов при различной структуре матрицы ковариации</p>	10	1

4	<p>Тема 4 Деревья решений. Основные понятия. Классы решаемых задач: описание данных, классификация, регрессия.</p> <p>Общий алгоритм построения дерева решений. Критерии выбора наилучшего атрибута: прирост информации, относительный прирост информации, индекс Гини. Правила останки разбиения дерева. Обрезание дерева. Алгоритм ID3. Переобучение деревьев решений. Обработка непрерывных атрибутов. Обучение на данных с пропусками. Программное обеспечение для построения деревьев решений</p>	6	1
5	<p>Тема 5 Алгоритмы сортировки. Задача сортировки. Внешняя и внутренняя сортировки. Сортировка подсчетом, простым включением, методом Шелла, простым извлечением, методом пузырька. Быстрая сортировка Хоара. Процедура разделения. Рекурсивный и не рекурсивный алгоритмы быстрой сортировки. Анализ сложности. Пирамидальная сортировка: турнирная сортировка, построение пирамиды и полное упорядочение. Анализ сложности алгоритма. Распределяющая (поразрядная) сортировка. Сравнение алгоритмов и программ внутренней сортировки. Внешняя сортировка. Порядковые статистики.</p>	4	2
ВСЕГО		34	8

4.2 Содержание практических занятий

№ п/п	№ раздела дисциплины	Наименование практических занятий	Объем в часах	
			О	З
1	2	Распознавание спамовых писем с помощью деревьев решений	10	4
2	3	Распознавание рукописных цифр с помощью наивного байесовского	16	4

3	5	Анализ многомерных данных. Корреляционные и причинноследственные связи. Корреляция признаков и структура данных. Латентные структуры в данных. Формальная и эффективная размерность данных. Структура и шум в данных. Понижение размерности данных. Поиск латентных структур. Отделение структуры от шума. Метод главных компонент как декомпозиция матрицы данных. Матрица счетов. Матрица нагрузок. Матрица ошибок. Объясненная и остаточная вариация в данных. Предобработка данных. Графическая интерпретация метода главных компонент. Критерии выбора количества главных компонент	4	2
4	5	Сегментация базы данных клиентов методами кластеризации и предсказание реакции клиента	2	2
ВСЕГО			32	12

4.3 Содержание лабораторных занятий

Учебным планом не предусмотрено

5 ПЕРЕЧЕНЬ ИННОВАЦИОННЫХ ФОРМ УЧЕБНЫХ ЗАНЯТИЙ

Преподавание дисциплины базируется на результатах научных исследований, проводимых УрТИСИ СибГУТИ, в том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей.

№ п/ п	Тема	Объем в часах		Вид учебных занятий	Используемые инновационные формы занятий
		О	З		
1	Распознавание рукописных цифр с помощью наивного байесовского	6	3	– практическое занятие;	– разбор конкретных ситуаций; – дискуссия;
2	Линейные структуры данных: связный список, стек, очередь, дек	6	3	– практическое занятие;	– разбор конкретных ситуаций; – дискуссия;
3	Основные методы машинного обучения	6	3	– лекционное занятие;	– разбор конкретных ситуаций; – дискуссия;
4	Сегментация базы данных клиентов методами кластеризации и предсказание реакции клиента	6	3	– практическое занятие;	– разбор конкретных ситуаций; – дискуссия;
ВСЕГО		2	12		

6 УЧЕБНО–МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ПО ДИСЦИПЛИНЕ

6.1 Список основной литературы

1. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие / А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/90591.html>. — Режим доступа: для авторизир. пользователей

6.2 Список дополнительной литературы

1. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/45418.html>. — Режим доступа: для авторизир. пользователей

6.3 Информационное обеспечение (в т.ч. интернет–ресурсы)

1. Официальный сайт UISI.RU/
2. Единая научно-образовательная электронная среда (Е-НОЭС) УрТИСИ <http://aur.uisi.ru/>
3. Электронная библиотечная система «IPRbooks» /<http://www.iprbookshop.ru/> доступ по логину и паролю
4. Электронный каталог АБК ASBOOK
5. Полнотекстовая база данных учебных и методических пособий СибГУТИ http://ellib.sibsutis.ru/cgi-bin/irbis64r_12/cgiirbis_64.exe?LNG=&C21COM=F&I21DBN=ELLIB&P21DBN=ELLIB&S21FMT=&S21ALL=&Z21ID=&S21CNR= доступ по логину и паролю
6. Электронные полнотекстовые издания ПГУТИ. http://ellib.sibsutis.ru/cgi-bin/irbis64r_12/cgiirbis_64.exe?LNG=&C21COM=F&I21DBN=PGUTI&P21DBN=PGUTI&S21FMT=&S21ALL=&Z21ID=&S21CNR= - доступ по паролю
7. Научная электронная библиотека (НЭБ) eLibrary <http://www.elibrary.ru>
8. Единое окно доступа к образовательным ресурсам <http://window.edu.ru/>

7 МАТЕРИАЛЬНО–ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Наименование аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Лекционная аудитория	Лекционные занятия	– компьютер; – мультимедийный проектор; – экран; – доска.
Компьютерный класс	практические занятия и самостоятельная работа	- персональные компьютеры, работающие под управлением операционной системы семейства Microsoft Windows, включенными в единую локальную сеть с выходом в Интернет;
Компьютерный класс	практические занятия	
Помещение для самостоятельной работы	самостоятельная работа	

8 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

8.1 Подготовка к лекциям

На лекциях необходимо вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание научных явлений и процессов, научные выводы и практические рекомендации.

Конспектирование лекций – сложный вид аудиторной работы, предполагающий интенсивную умственную деятельность студента. Целесообразно сначала понять основную мысль, излагаемую лектором, а затем записать ее. Желательно оставлять поля, на которых при самостоятельной работе с конспектом можно сделать дополнительные записи и отметить непонятные вопросы.

Конспект лекции лучше подразделять на пункты в соответствии с вопросами плана лекции, предложенными преподавателем. Следует обращать внимание на акценты, выводы, которые делает лектор, отмечая наиболее важные моменты в лекционном материале.

Во время лекции можно задавать преподавателю уточняющие вопросы с целью освоения теоретических положений, разрешения спорных вопросов.

8.2 Подготовка к практическим занятиям

Подготовку к практическим занятиям следует начинать с ознакомления плана практического занятия, который отражает содержание предложенной темы. Изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучении основной и дополнительной литературы. Новые понятия по изучаемой теме необходимо выучить 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 практикума, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильно выполнении практических заданий и контрольных работ.

Целесообразно начать с изучения основной литературы в части учебников и учебных пособий. Далее рекомендуется перейти к анализу научных монографий и статей, рассматривающих отдельные аспекты проблем, изучаемых в рамках дисциплины, а также официальных Интернет–ресурсов, в которых могут содержаться основные вопросы изучаемой проблемы.

При работе с литературой необходимо:

- сопоставлять, сравнивать, классифицировать, группировать, систематизировать информацию с определенной учебной задачей;
- обобщать полученную информацию, оценивать прослушанное и прочитанное;
- фиксировать основное содержание сообщений; формулировать устно и письменно, основную идею сообщения; составлять план, формулировать тезисы;
- готовить доклады и презентации к ним;
- работать в разных режимах (индивидуально, в паре в группе) взаимодействуя друг с другом;
- пользоваться реферативными и справочными материалами;

– обращаться за помощью, дополнительными разъяснениями к преподавателю, другим студентам.

– пользоваться словарями и др.

8.3 Самостоятельная работа студентов

Успешное освоение компетенций, формируемых данной учебной дисциплиной, предполагает оптимальное использование времени самостоятельной работы.

Подготовка к лекционным занятиям включает выполнение всех видов заданий, рекомендованных к каждой лекции, т.е. задания выполняются еще до лекционного занятия по соответствующей теме. Целесообразно дорабатывать свой конспект лекции, делая в нем соответствующие записи из литературы, рекомендованной преподавателем и предусмотренной учебной программой.

Все задания к практическим занятиям, а также задания, вынесенные на самостоятельную работу, рекомендуется выполнять непосредственно после соответствующей темы лекционного курса, что способствует усвоению материала, позволяет своевременно выявить и устранить «пробелы» в знаниях, систематизировать ранее пройденный материал, на его основе приступить к получению новых знаний и овладению навыками.

Самостоятельная работа во внеаудиторное время состоит из:

- повторение лекционного материала;
- подготовки к практическим занятиям;
- изучения учебно-методической и научной литературы;
- решения задач, выданных на практических занятиях;
- подготовки к контрольным работам, тестированию и т.д.;
- подготовки рефератов и иных индивидуальных письменных работ по заданию преподавателя;
- проведение самоконтроля путем ответов на вопросы текущего контроля знаний, решения представленных в учебно-методических материалах дисциплины задач, тестов, написания рефератов и эссе по отдельным вопросам изучаемой темы.

8.4 Подготовка к промежуточной аттестации

При подготовке к промежуточной аттестации необходимо:

- внимательно изучить перечень вопросов и определить, в каких источниках находятся сведения, необходимые для ответа на них;
- внимательно прочитать рекомендуемую литературу;
- изучить слайды по темам дисциплины «Методы машинного обучения»;
- составлять краткие конспекты ответов (планы ответов).

Текущий контроль достижения результатов обучения по дисциплине включает следующие процедуры:

- решение индивидуальных задач на практических занятиях;
- контроль самостоятельной работы, осуществляемый на каждом практическом занятии;
- защита курсового проекта.

Промежуточный контроль достижения результатов обучения по дисциплине проводится в следующих формах:

Для проведения текущего контроля и промежуточной аттестации используются оценочные средства, описание которых расположено в Приложении 1 и на сайте (<http://www.aup.uisi.ru>).