

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

УТВЕРЖДАЮ
Директор по образовательной деятельности

С.Т. Князев
2021 г.



РАБОЧАЯ ПРОГРАММА МОДУЛЯ

Код модуля
М.1.6

Модуль
Машинное обучение

Екатеринбург, 2021

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Прикладной анализ данных	Код ОП 09.04.02
Направление подготовки Информационные системы и технологии	Код направления и уровня подготовки 09.04.02

Области образования, в рамках которых реализуется модуль образовательной программы по СУОС УрФУ:

№ п/п	Перечень областей образования, для которых разработан СУОС УрФУ	Уровень подготовки
1.	Инженерное дело, технологии и технические науки	магистратура

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Долганов Антон Юрьевич	кандидат технических наук, без учёного звания	Доцент	Кафедра радиоэлектроники и телекоммуникаций, ИРИТ-РТФ, УрФУ
2	Патракова Екатерина Сергеевна	-	Ведущий исследователь данных	ПАО Сбербанк

Рекомендовано учебно-методическим советом института радиоэлектроники и информационных технологий - РТФ

Протокол № 4 от 28.05.2021 г.

Согласовано:

Управление образовательных программ



Р.Х.Токарева

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Машинное обучение

1.1. Аннотация содержания модуля

Модуль «Машинное обучение» состоит из одноименной дисциплины.

Содержание дисциплин модуля позволит студентам получить комплексное всестороннее представление об основных методах машинного обучения. Сформировать у студентов понимание алгоритмов машинного обучения, инфраструктуры Big Data и технологий искусственного интеллекта. Студенты смогут освоить основные подходы и получить навыки решения практических задач.

1.2. Структура и объем модуля

Таблица 1

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах
1.	Машинное обучение	12 з.е. / 432 ч.
ИТОГО по модулю:		12 з.е. / 432 ч.

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	<i>отсутствуют</i>
Постреквизиты и корреквизиты модуля	<i>отсутствуют</i>

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Планируемые результаты обучения (индикаторы)
1	2	3
Машинное обучение	ОПК-2 - Способен самостоятельно ставить, формализовывать и решать задачи, относящиеся к профессиональной деятельности, используя методы моделирования и математического анализа	З-1 - Сделать обзор основных методов моделирования и математического анализа, применимых для формализации и решения задач профессиональной деятельности З-2 - Характеризовать сферы применения и возможности пакетов прикладных программ для решения задач

		<p>профессиональной деятельности</p> <p>У-1 - Самостоятельно сформулировать задачу области профессиональной деятельности, решение которой требует использования методов моделирования и математического анализа</p> <p>У-2 - Использовать методы моделирования и математического анализа, в том числе с использованием пакетов прикладных программ для решения задач профессиональной деятельности</p> <p>П-1 - Решать самостоятельно сформулированные практические задачи, относящиеся к профессиональной деятельности методами моделирования и математического анализа, в том числе с использованием пакетов прикладных программ</p> <p>Д-1 - Проявлять ответственность и настойчивость в достижении цели</p>
	<p>ОПК-5 - Способен планировать, организовывать и контролировать работы по созданию, установке и модернизации технологического оборудования и технологических процессов в сфере своей профессиональной деятельности</p>	<p>З-4 - Показать возможности использования цифровых технологий (создание цифровых двойников) для оптимизации работы по созданию, установке и модернизации технологического оборудования, технологических процессов и информационных систем</p> <p>У-4 - Использовать при необходимости техники цифрового моделирования при выполнении работ по созданию, установке и модернизации технологического оборудования, технологических процессов и информационных систем</p> <p>П-1 - Самостоятельно составить план работ в целом по этапам создания, установки и модернизации технологического оборудования, технологических процессов и информационных систем либо отдельных этапов этой работы</p>

	<p>ПК-2 - Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений</p>	<p>З-1 - Интерпретировать результаты применения математических моделей процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений</p> <p>З-3 - Сделать обзор методов математического моделирования и анализа данных</p> <p>У-1 - Определять оптимальные математические модели для решения задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений в зависимости от заданных условий</p> <p>П-1 - Иметь практический опыт применения математического моделирования и анализа данных</p>
--	--	--

1.5. Форма обучения

Обучение по дисциплине модуля может осуществляться в очной форме.

2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН МОДУЛЯ

**ПРОГРАММА МОДУЛЯ
МАШИННОЕ ОБУЧЕНИЕ**

**РАЗДЕЛ 2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН
МОДУЛЯ**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
МАШИННОЕ ОБУЧЕНИЕ**

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Долганов Антон Юрьевич	кандидат технических наук, без учёного звания	Доцент	Кафедра радиоэлектроники и телекоммуникаций, ИРИТ-РТФ, УрФУ
2	Патракова Екатерина Сергеевна	-	Ведущий исследователь данных	ПАО Сбербанк

Рекомендовано учебно-методическим советом института радиоэлектроники и информационных технологий - РТФ

Протокол № 4 от 28.05.2021 г.

2. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ Машинное обучение

2.1. Технологии обучения, используемые при изучении дисциплины модуля

Традиционная (репродуктивная) технология

2.2. Содержание дисциплины

Таблица 1

Код раздела, темы	Раздел, тема дисциплины	Содержание
1.	Задачи машинного обучения	Машинное обучение. Этапы разработки ds проекта. Задачи классического машинного обучения
2.	Линейные алгоритмы	Алгоритмы линейного машинного обучения: линейная регрессия, логистическая регрессия и линейный дискриминантный анализ. Регуляризация L1 и L2.
3.	White-box алгоритмы классификации	Простые алгоритмы классификации (Naive Bayse, SVM, Decision Tree). Плюсы и минусы алгоритмов классификации. Оценка качества алгоритмов. Отбор признаков.
4.	Многоклассовая классификация	Многоклассовая классификация. Стратегии обучения one vs rest, one vs all. Метрики для многоклассовой классификации. Macro и micro усреднение.
5.	К ближайших соседей	К ближайших соседей, метрики расстояния, преимущества и недостатки алгоритма. Улучшения алгоритма.
6.	Ансамбли	Ансамбли: корректирующие операции, стэкинг, блендинг, беггинг (на примере Random Forest), бустинг (адаптивный, градиентный, xgboost). Градиентный спуск.
7.	Алгоритмы кластеризации	Кластеризация: kmeans, DBSCAN, Агломеративная кластеризация. Оценка качества кластеризации.
8.	Алгоритмы обработки естественного языка	Введение в Nature Language Processing. Типы задач. Сложности в NLP. Предобработка текстов. fasttext, elmo. Визуализация эмбедингов. Практика: Предобработка текстов. Векторизация текстов и эмбединги на задаче классификации.
9.	Векторизация текстов	Векторизация текстов: one-hot-encoding, labael encoding, bag-of-words, tf-idf. Плюсы и минусы подходов.
10.	Эммбединги	Эммбединги: word2vec, fasttext, elmo, BERT. Архитектуры сетей и принцип их обучения. Skip-gram и Continuous Bag Of Words. Использование предобученных эмбедингов.

11.	Бинарный поиск для текстов	Методы бинарного поиска faiss, nmslib на эмбедингах. Метрики похожести текстов.
12.	Введение в нейронные сети	Введение в нейронные сети. Простые архитектуры на полносвязных сетях. Функции активаций. Оптимизаторы.
13.	Основы программирования нейронных сетей	Библиотеки для обучения нейронных сетей. Основы TensorFlow. Написание первой нейронной сети.
14.	Нейронные сети для задачи анализа изображений	Сверточные нейронные сети. Распознавание объектов на изображении. Предварительно обученные нейронные сети. Проблема исчезающих градиентов. Регуляризации.
15.	Нейронные сети для задачи анализа текстов	Нейронные сети для задач обработки естественного языка. Одномерные сверточные нейронные сети WordCNN. CharCNN.
15.	Рекуррентные нейронные сети для задач анализа текстов	Рекуррентные нейронные сети. GRU, LSTM. Vanilla RNN. Backpropagation through time.
16.	Задача морфологического анализа текста.	Морфологический анализ. Задачи тегирования частей речи, падежей.
17.	Задача выделения сущностей из текстов	Задача выделения сущностей из текстов. Разметка данных для решения задачи BIO. Архитектуры нейронных сетей для задачи NER.
18.	Задача синтаксического анализа текста	Задача синтаксического анализа текста. Разметка Universal Dependencies. Библиотеки для синтаксического анализа UD Pipe. Применение предобученных моделей для разных языков.
19.	Чат-боты	Чат-боты: понятие интента, triplet loss, заполнение слотов.
20.	Seq2seq архитектуры для генерации текстов	Seq2seq и Attention. Transfer learning и Fine tuning.

2.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

2.4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ Машинное обучение

Электронные ресурсы (издания)

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. URL: <http://www.machinelearning.ru/> .
2. Цитульский Антон Максимович, Иванников Александр Владимирович, Рогов Илья Сергеевич NLP - Обработка естественных языков // StudNet. 2020. №6. URL: <https://cyberleninka.ru/article/n/nlp-obrabotka-estestvennyh-yazykov> .
3. Браславский П.И. Введение в обработку естественного языка. URL: <https://stepik.org/course/1233/>.

Профессиональные базы данных, информационно-справочные системы

1. Цифровая библиотека научно-технических изданий Института инженеров по электротехнике и радиоэлектронике (Institute of Electrical and Electronic Engineers (IEEE)) на английском языке – <http://www.ieee.org/ieeexplore>
2. Oxford University Press – <http://www.oxfordjournals.org/en/>
3. Архив препринтов с открытым доступом – <https://arxiv.org/>

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

1. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
2. eBook Collections Springer Nature – <https://link.springer.com/>
3. Гугл Академия – <https://scholar.google.ru/>
4. Электронный научный архив УрФУ <https://elar.urfu.ru/>
5. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
6. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
7. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
8. Университетская библиотека ONLINE – <https://biblioclub.ru/>
9. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
10. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
11. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

2.5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Машинное обучение

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 2

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
1.	Лекции	Мультимедийный проектор с экраном.	Используется бесплатно-распространяемое программное обеспечение: 1. Python – https://www.python.org/ 2. Веб - среда разработки для языка программирования Python: Kaggle - https://www.kaggle.com/
2.	Лабораторные занятия	Компьютерный класс. Мультимедийный проектор с экраном; Сетевое оборудование; Локальная сеть с выходом в глобальную сеть Интернет.	