

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
«10» 25 апреля 2022

РАБОЧАЯ ПРОГРАММА МОДУЛЯ



Код модуля	Модуль
1161158	Приложения машинного обучения

Екатеринбург

2022

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Алгоритмы искусственного интеллекта	Код ОП 09.03.01
Направление подготовки Информатика и вычислительная техника	Код направления и уровня подготовки 09.03.01

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	бакалавриат

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Новиков Максим Юрьевич	Кандидат педагогических наук, без ученого звания	Доцент	Кафедра «Аналитика больших данных и методы видеоанализа»
2	Созыкин Андрей Владимирович	Кандидат технических наук, без ученого звания	Доцент	Кафедра информационных технологий и систем управления

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Приложения машинного обучения

1.1. Аннотация содержания модуля

Модуль «Приложения машинного обучения» нацелен на формирование у будущих специалистов понимания разработки программных решений в области машинного обучения на современных языках программирования. В рамках дисциплины студенты научатся решать задачи машинного обучения, применяя один или несколько технологических стеков, познакомятся с прикладными задачами искусственного интеллекта и определяют границы применимости языков программирования и программно-аппаратных решений.

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
1	Приложения машинного обучения	3
ИТОГО по модулю:		3

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	Не предусмотрены
Постреквизиты и кореквизиты модуля	Не предусмотрены

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

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
1	2	3	4
Приложения машинного обучения	ПК-2. Способен разрабатывать и тестировать программные компоненты решения задач в системах искусственного интеллекта	ПК-2.1. Разрабатывает приложения систем искусственного интеллекта	ПК-2.1. 3-1. Знает современные языки программирования, библиотеки и программные платформы для функционального, логического, объектно-ориентированного программирования приложений систем искусственного интеллекта (Python, R, C++, CC)

			ПК-2.1. У-1. Умеет разрабатывать программные приложения систем искусственного интеллекта, с использованием современных языков программирования, библиотек и программных платформ функционального, логического, объектно-ориентированного программирования (Python, R, C++, Cd)
	ПК-4. Способен использовать инструментальные средства для решения задач машинного обучения	ПК-4.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	ПК-4.1. З-1. Знает возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения ПК-4.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения

1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной форме.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Приложения машинного обучения

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Новиков Максим Юрьевич	Кандидат педагогических наук, без ученого звания	Доцент	Кафедра «Аналитика больших данных и методы видеоанализа»
2	Созыкин Андрей Владимирович	Кандидат технических наук, без ученого звания	Доцент	Кафедра информационных технологий и систем управления

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - Базовый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания; Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

1.2. Содержание дисциплины

Таблица 1.1

Код раздел, темы	Раздел, тема дисциплины*	Содержание
1	Введение	Жизненный цикл аналитики данных. Исследовательский анализ данных. Категории задач машинного обучения. Подготовка данных.
2	Методы кластеризации и понижения размерности	Понижение размерности - метод главных компонент (PCA). Отбор числа главных компонент. Метод понижения размерности t-SNE. Методы кластеризации. Алгоритм k-means. Выбор числа кластеров для kMeans.. Метрики качества кластеризации. Визуализация результатов кластеризации.
3	Классификация данных	Деревья решений в Scikit-learn. Метод ближайших соседей (kNN). Estimator API в Scikit-learn. Кроссвалидация. Ансамбли моделей. Отбор признаков и параметров модели с помощью GridSearch. Виды классификаторов Naive Bayes в Scikit-learn. Настройка параметров Naive Bayes. Naive Bayes в ситуации дисбаланса классов. Бутстрэп. Бэггинг. Out-of-bag error. Параметры алгоритма Random Forest. Подбор и оценка влияния параметров Random Forest. Важность признаков в Random Forest. Основные метрики качества классификации. Характеристики бинарного классификатора. Многоклассовая классификация.LASSP и Ridge-регрессия. Полиномиальные преобразования переменных для задачи регрессии.
4	Машинное обучение в прикладных задачах	Анализ временных рядов и прогнозирование численных признаков. Операционализация моделей машинного обучения. Парсинг данных с помощью метода Pandas read_html. Работа с библиотеками Beautiful Soup и Selenium. Работа с API-интерфейсами веб-сервисов. Построение социального графа. Графовые представления данных. Тематическое моделирование отзывов с сайта banki.ru с помощью метода LDA. Настройка параметров LDA. Визуализация тематических моделей. Задачи сентимент-анализа. Подготовка данных для сентимент-анализа.

		Модель мешка слов. Логистическая регрессия и SVM в задачах классификации текстов. Визуализация термов и влияния на результат классификации. Настройка параметров моделей классификации текстов.
--	--	---

1.3. Направление, виды воспитательной деятельности и используемые технологии

Таблица 1.2

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения
Профессиональное воспитание	профориентационная деятельность	Технология формирования уверенности и готовности к самостоятельной успешной профессиональной деятельности Технология самостоятельной работы	ПК-2. Способен разрабатывать и тестировать программные компоненты решения задач в системах искусственного интеллекта	ПК-2.1. 3-1. Знает современные языки программирования, библиотеки и программные платформы для функционального, логического, объектно-ориентированного программирования приложений систем искусственного интеллекта (Python, R, C++, CC)

1.4. Программа дисциплины реализуется на государственном языке Российской Федерации .

1.1.1.1. 2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

1.1.1.2. Приложения машинного обучения

Электронные ресурсы (издания)

1. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск: УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053>. — Режим доступа: для авториз. пользователей.
2. Мейер, Б. Инструменты, алгоритмы и структуры данных / Б. Мейер. - 2-е изд., испр. - Москва: Национальный Открытый Университет «ИНТУИТ», 2016. - 543 с.: схем., ил. - Библиогр. в кн. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=429033>.
3. Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения: руководство / С. Рашка; перевод с английского А. В. Логунова. — Москва: ДМК Пресс, 2017. — 418 с. — ISBN 978-5-97060-409-0. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/100905>. — Режим доступа: для авториз. пользователей.

4. Северенс, Ч. Введение в программирование на Python / Ч. Северенс. - 2-е изд., испр. - Москва: Национальный Открытый Университет «ИНТУИТ», 2016. - 231 с.: схем., ил.; [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=429184>.
5. Фарунцев С. Д.; Интеллектуальные технологии управления в технических системах: учебное пособие; Омский государственный технический университет (ОмГТУ), 2019; http://biblioclub.ru/index.php?page=book_red&id=682099 (Электронное издание)
6. Воронцов К. В. Машинное обучение Школа Анализа данных Яндекс. МФТИ. национальный открытый университет Интуит, 2015. – Режим доступа: свободный – URL: <https://www.intuit.ru/studies/courses/13844/1241/info>.
7. Шапиро Л. , Стокман Д.; Компьютерное зрение: учебное пособие; Москва: Лаборатория знаний, 2020; <https://biblioclub.ru/index.php?page=book&id=445962> (Электронное издание)
8. Сурова Н. Ю. , Косов М. Е.; Искусственный интеллект: монография; Москва: Юнити-Дана, 2021; http://biblioclub.ru/index.php?page=book_red&id=690578 (Электронное издание)
9. Рыбина Г. В.; Основы построения интеллектуальных систем: учебное пособие; Москва: Финансы и статистика, 2021; http://biblioclub.ru/index.php?page=book_red&id=685190 (Электронное издание)
10. Джонс М. Т.; Программирование искусственного интеллекта в приложениях: практическое пособие; Москва: ДМК Пресс, 2018; http://biblioclub.ru/index.php?page=book_red&id=686757 (Электронное издание)
11. Сырецкий Г. А.; Искусственный интеллект и основы теории интеллектуального управления: практикум : в 3 частях, Ч. 1. Фазисистемы; Новосибирск: Новосибирский государственный технический университет, 2016; http://biblioclub.ru/index.php?page=book_red&id=576318 (Электронное издание)
12. Харахан О. Г.; Системы искусственного интеллекта : Практикум для проведения лабораторных работ: учебное пособие, Ч. 1; Москва: Московский государственный горный университет, 2006; http://biblioclub.ru/index.php?page=book_red&id=83722 (Электронное издание)
13. Крейман Г.; Биологическое и компьютерное зрение: научно-популярное издание; Москва: ДМК Пресс, 2022; http://biblioclub.ru/index.php?page=book_red&id=694657 (Электронное издание)
14. Брокшмидт К.; Введение в разработку приложений для Windows 8 с использованием HTML, CSS и JavaScript: курс лекций; Москва: Национальный Открытый Университет «ИНТУИТ», 2016; http://biblioclub.ru/index.php?page=book_red&id=428973 (Электронное издание)

Профессиональные базы данных, информационно-справочные системы

1. Единое окно доступа к образовательным ресурсам. Раздел Информатика и информационные технологии <http://window.edu.ru/catalog>
2. Интернет-Университет Информационных Технологий <http://www.intuit.ru/>
3. Веб-сервис для хостинга IT-проектов и их совместной разработки Github <http://www.github.ru>

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. ЭБС Университетская библиотека онлайн «Директ-Медиа» <http://www.biblioclub.ru/>

2. eLibrary ООО Научная электронная библиотека <http://elibrary.ru>

1.1.1.3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

1.1.1.4. Приложения машинного обучения

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения
1	Лекции	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Подключение к сети Интернет	Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Office 365 EDUA3 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES
2	Лабораторные занятия	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Периферийное устройство Персональные компьютеры по количеству обучающихся Оборудование, соответствующее требованиям организации учебного процесса в соответствии с санитарными правилами и нормами Подключение к сети Интернет	Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES Python Software
3	Консультации	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Персональные компьютеры по количеству обучающихся Оборудование, соответствующее требованиям организации учебного процесса в соответствии с санитарными правилами и нормами	Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES

		Подключение к сети Интернет	
4	Текущий контроль и промежуточная аттестация	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Оборудование, соответствующее требованиям организации учебного процесса в соответствии с санитарными правилами и нормами</p> <p>Подключение к сети Интернет</p>	<p>Microsoft Windows 8.1 Pro 64-bit</p> <p>RUS OLP NL Acdmc</p> <p>Office 365 EDUA5 ShrdSvr ALNG</p> <p>SubsVL MVL PerUsr B Faculty EES</p> <p>Python Software</p>
5	Самостоятельная работа студентов	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Оборудование, соответствующее требованиям организации учебного процесса в соответствии с санитарными правилами и нормами</p> <p>Подключение к сети Интернет</p>	<p>Microsoft Windows 8.1 Pro 64-bit</p> <p>RUS OLP NL Acdmc</p> <p>Office 365 EDUA5 ShrdSvr ALNG</p> <p>SubsVL MVL PerUsr B Faculty EES</p> <p>Python Software</p>

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Приложения машинного обучения

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Новиков Максим Юрьевич	Кандидат педагогических наук, без ученого звания	Доцент	Кафедра «Аналитика больших данных и методы видеоанализа»
2	Созыкин Андрей Владимирович	Кандидат технических наук, без ученого звания	Доцент	Кафедра информационных технологий и систем управления

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Приложения машинного обучения

2.	Объем дисциплины в зачетных единицах	3	
3.	Виды аудиторных занятий	Лекции Лабораторные занятия	
4.	Промежуточная аттестация	Зачет	
5.	Текущая аттестация	Домашняя работа	2

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Приложения машинного обучения

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1		2	3
ПК-2. Способен разрабатывать и тестировать программные компоненты решения задач в системах искусственного интеллекта	ПК-2.1. Разрабатывает приложения систем искусственного интеллекта	ПК-2.1. Знает современные языки программирования, библиотеки и программные платформы для функционального, логического, объектно-ориентированного программирования приложений систем искусственного интеллекта (Python, R, C++, CC) ПК-2.1. Умеет разрабатывать программные приложения систем искусственного интеллекта, с использованием современных языков программирования, библиотек и программных платформ функционального, логического, объектно-ориентированного	Домашняя работа №1 Домашняя работа №2 Лабораторные занятия Лекции Зачёт

		программирования (Python, R, C++, Cd)	
ПК-4. Способен использовать инструментальные средства для решения задач машинного обучения	ПК-4.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	ПК-4.1. З-1. Знает возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения ПК-4.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения	Контрольная работа Домашняя работа Лабораторные занятия Лекции Зачёт Экзамен

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Домашняя работа №1	4, 8	50
Домашняя работа №2	4, 16	50
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – зачет		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – не предусмотрено		

3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий –0.5		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>отчет по лабораторным работам</i>	4, 1-16	100
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям -1		
Промежуточная аттестация по лабораторным занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Лабораторные занятия

Примерный перечень тем:

1. Базовая очистка и подготовка данных
2. Визуализация данных средствами Python
3. Метод кластеризации и понижения размерности
4. Методы классификации
5. Регрессионные модели в Python
6. Анализ временных рядов и прогнозирование численных признаков средствами Python

7. Анализ неструктурированных данных
8. Операционализация моделей машинного обучения
9. Методы Web-Scraping
10. Методы построения социальных графов
11. Методы тематического моделирования
12. Оценка эмоциональной окраски

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

5.2.1. Домашняя работа №1

«Логистическая регрессия и случайный лес в задаче кредитного скоринга»

1. Ознакомьтесь с содержанием демонстрационных блокнотов
2. Создайте новый блокнот, импортируйте необходимые библиотеки
3. Задача – построить модель для задачи кредитного скоринга
4. Выполните следующие блоки заданий
 - Сделайте интервальную оценку среднего возраста для клиентов, которые просрочили выплату кредита с 90% уверенностью..
 - Сделайте GridSearch с метрикой "roc-auc" по параметру C. Какое оптимальное значение параметра C?
 - Можно ли считать лучшую модель устойчивой? (модель считаем устойчивой, если стандартное отклонение на валидации меньше 0.5%)
 - Определите самый важный признак. Важность признака определяется абсолютным значением его коэффициента. Или используя метод SelectKBest
 - Оцените логистическую регрессию в абсолютных величинах. После этого посчитайте во сколько раз увеличатся шансы, что клиент не выплатит кредит, если увеличить возраст на 20 лет при всех остальных равных значениях признаков.
 - На сколько точность лучшей модели случайного леса выше точности логистической регрессии на валидации?
 - Определите какой признак имеет самое слабое влияние (используя методы оценки важности признаков в Random Forest)
 - Какое наиболее существенное преимущество логистической регрессии перед случайным лесом для нашей задачи?
 - обучить бэггинг классификатор (`random_state=42`). В качестве базовых классификаторов возьмите 100 логистических регрессий и на этот раз используйте не GridSearchCV, а RandomizedSearchCV. Так как перебирать все 54 варианта комбинаций долго, то поставьте максимальное число итераций 20 для RandomizedSearchCV. Также не забудьте передать параметр валидации cv и `random_state=1`. Какая лучшая точность получилась?
4. Вы должны загрузить `ipynb` вашего решения или ссылку (если дана ссылка, вам нужно убедиться, что режим доступа открыт)

5.2.2. Домашняя работа №2

«Линейная регрессия, RF-регрессия в задаче по определению качества вина»

1. Ознакомьтесь с содержанием демонстрационных блокнотов
2. Создайте новый блокнот, импортируйте необходимые библиотеки
 - Каковы среднеквадратичные ошибки линейной регрессии на обучающей и отложенной выборках?
 - Какой признак линейная регрессия считает наиболее сильно влияющим на качество вина?

- Каковы среднеквадратичные ошибки случайного леса на обучающей выборке, на кросс-валидации (`cross_val_score` с `scoring='neg_mean_squared_error'` и остальными параметрами по умолчанию) и на отложенной выборке?
 - Обучите лес с параметрами `max_depth=19`, `max_features=7`, и `min_samples_leaf=1`. Каковы среднеквадратичные ошибки настроенного случайного леса на обучающей выборке, на кросс-валидации (`cross_val_score` с `scoring='neg_mean_squared_error'`) и на отложенной выборке?
 - Какой признак оказался главным в настроенной модели случайного леса?
 - Сделайте выводы о качестве моделей и оценках влияния признаков на качество вина.
4. Вы должны загрузить `ipynb` вашего решения или ссылку (если дана ссылка, вам нужно убедиться, что режим доступа открыт)

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Зачет

1. Жизненный цикл аналитики данных
2. Категории задач машинного обучения
3. Методы кластеризации и понижения размерности
4. Деревья решений в Scikit-learn.
5. Метод ближайших соседей (kNN).
6. Estimator API в Scikit-learn.
7. Кроссвалидация.
8. Ансамбли моделей.
9. Отбор признаков и параметров модели с помощью GridSearch.
10. Виды классификаторов Naive Bayes в Scikit-learn. Настройка параметров Naive Bayes. Naive Bayes в ситуации дисбаланса классов.
11. Бутстрэп.
12. Бэггинг. Out-of-bag error.
13. Параметры алгоритма Random Forest. Подбор и оценка влияния параметров Random Forest. Важность признаков в Random Forest.
14. Основные метрики качества классификации. Характеристики бинарного классификатора.
15. Многоклассовая классификация. LASSP и Ridge-регрессия. Полиномиальные преобразования переменных для задачи регрессии.
16. Анализ временных рядов и прогнозирование численных признаков.
17. Операционализация моделей машинного обучения.
18. Графовые представления данных. Визуализация тематических моделей.
19. Задачи сентимент-анализа. Подготовка данных для сентимент-анализа. Модель мешка слов.
20. Логистическая регрессия и SVM в задачах классификации текстов. Визуализация термов и влияния на результат классификации. Настройка параметров моделей классификации текстов.