

Документ подписан простой электронной подписью
Информационный сертификат:
ФИО: Клочков Юрий Сергеевич
Должность: и.о. ректора
Дата подписания: 26.04.2024 12:44:54
Уникальный программный ключ:
4e7c4ea90328ec8e65c5d8058549a2538d7400d1

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Председатель КСН


О.М. Барбаков

« 27 » мая 2021г.

РАБОЧАЯ ПРОГРАММА


дисциплины:	Анализ данных и машинное обучение
направление подготовки:	38.03.05 Бизнес - информатика
направленность:	Информационные системы предприятия
форма обучения:	очная

Рабочая программа разработана в соответствии с утвержденным учебным планом от 27.05.2021г. и требованиями ОПОП ВО по направлению подготовки 38.03.05 Бизнес – информатика, направленность Информационные системы предприятия к результатам освоения дисциплины «Анализ данных и машинное обучение».

Рабочая программа рассмотрена
на заседании кафедры бизнес – информатики и математики


Протокол № 12 от « 27 » мая 2021г.

Заведующий кафедрой БИМ


_____ О.М. Барбаков

СОГЛАСОВАНО:


Заведующий выпускающей кафедрой


_____ О.М. Барбаков


« 27 » мая 2021г.

Рабочую программу разработал:

М.А. Аханова, доцент, к.с.н.



С.В. Овчинникова, доцент, к.с.н.



1. Цели и задачи освоения дисциплины

Цель дисциплины: освоение фундаментальных понятий анализа данных, технологий и алгоритмов машинного обучения, используемых для интеллектуального анализа данных.

Задачи дисциплины:

- овладение основными понятиями и принципами анализа данных;
- изучение современных технологий и алгоритмов интеллектуального анализа данных, в том числе алгоритмов машинного обучения;
- формирование практических навыков использования технологий и алгоритмов интеллектуального анализа данных;
- формирование первичных навыков самостоятельной разработки алгоритмов интеллектуального анализа данных и машинного обучения для решения практических задач;
- развитие у обучающихся творческого и интеллектуального потенциала.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина относится к дисциплинам обязательной части учебного плана.

Необходимыми условиями для освоения дисциплины являются:

знание:

- основ алгебры, математического анализа, теории вероятностей, математической статистики и случайных процессов;

умение:

- владеть навыками алгоритмизации и программирования;

владение:

- навыками использования компьютерных технологий и средств обработки информации.

Содержание дисциплины служит основой для научно-исследовательской работы, прохождения преддипломной практики, подготовки выпускной квалификационной работы.

3. Результаты обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Таблица 3.1

Код и наименование компетенции	Код и наименование индикатора достижения компетенции (ИДК)	Код и наименование результата обучения по дисциплине
ОПК – 4 Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее	ОПК – 4.1 Использует основные методы и средства сбора, обработки и анализа информации, в том числе работу с большими данными	Знать (З1) основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения
		Уметь (У1) использовать

сбора, обработки и анализа для информационно – аналитической поддержки принятия управленческих решений	математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения
	Владеть (В1) практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения

4. Объем дисциплины

Общий объем дисциплины составляет 4 зачетных единиц, 144 часов.

Таблица 4.1

Форма обучения	Курс/ семестр	Аудиторные занятия/контактная работа, час.			Самостоятельная работа, час.	Форма промежуточной аттестации
		Лекции	Практические занятия	Лабораторные занятия		
очная	4/7	14	-	42	88	Экзамен

5. Структура и содержание дисциплины

5.1. Структура дисциплины

- очная форма обучения (ОФО)

Таблица 5.1

№ п/п	Структура дисциплины		Аудиторные занятия, час.			СРС, час.	Всего, час.	Код ИДК	Оценочные средства
	Номер раздела	Наименование раздела	Л.	Пр.	Лаб.				
1	1	Введение. Технологии анализа данных	1	-	2	-	3	ОПК 4.1	Задание к лабораторной работе №1
2	2	Классификация и регрессия	1	-	6	4	11		Задание к лабораторной работе №2,3
3	3	Предсказательная аналитика на основе ансамблевого обучения	1	-	4	4	9		Задание к лабораторной работе №4
4	4	Распознавание образов	2	-	6	8	16		Задание к лабораторной работе №5,6
5	5	Генетические алгоритмы	1	-	4	4	9		Задание к лабораторной работе №7
6	6	Обработка естественного языка	2	-	6	10	18		Задание к лабораторной работе №8
7	7	Вероятностный подход к обработке последовательных данных	1	-	2	4	7		Задание к лабораторной работе №9
8	8	Распознавание речи	1	-	4	6	11		Вопросы к коллоквиуму
9	9	Обнаружение и	2	-	4	6	12		Вопросы к

		отслеживание объектов							коллоквиуму
10	10	Искусственные нейронные сети	2	-	4	6	12		Задание к лабораторным работам №10
11	Экзамен		-	-	-	36	36		Вопросы для экзамена
Итого:			14	-	42	88	144	X	X

- заочная форма обучения (ЗФО): не реализуется
- очно-заочная форма обучения (ОЗФО): не реализуется

5.2. Содержание дисциплины.

5.2.1. Содержание разделов дисциплины (дидактические единицы)

Раздел 1. Введение. Технологии анализа данных.

Введение в анализ данных. Принципы анализа данных. Структурированные данные. Технологии KDD и Data mining. Происхождение и понимание термина «искусственный интеллект». Области применения и направления развития искусственного интеллекта. Data Mining, Machine Learning и Knowledge Discovery in Databases. Основные стандарты процесса KDD&DM.

Раздел 2. Классификация и регрессия.

Обучение с учителем и без учителя. Классификация. Предварительная обработка данных. Кодирование меток. Логистический классификатор. Наивный байесовский классификатор. Матрица неточностей. Машины опорных векторов. Регрессия. Создание регрессора одной переменной. Многомерный регрессор.

Раздел 3. Предсказательная аналитика на основе ансамблевого обучения.

Ансамблевое обучение. Построение моделей обучения посредством ансамблевого метода. Деревья принятия решений. Случайные и предельно случайные леса. Создание классификаторов на основе случайных и предельно случайных лесов. Оценка мер достоверности прогнозов. Обработка дисбаланса классов. Нахождение оптимальных обучающих параметров с помощью сеточного поиска.

Раздел 4. Распознавание образов.

Метод k-средних. Кластеризация данных с помощью метода k-средних. Оценка количества кластеров с использованием метода сдвига среднего. Оценка качества кластеризации с помощью силуэтных оценок. Смешанные гауссовские модели. Классификатор на основе гауссовской смешанной модели. Обучающий конвейер. Извлечение ближайших соседей. Создание классификатора методом k ближайших соседей. Вычисление оценок сходства.

Раздел 5. Генетические алгоритмы.

Эволюционные и генетические алгоритмы. Основные понятия генетических алгоритмов. Генерация битовых образов с предопределенными параметрами. Визуализация хода эволюции. Решение задачи символической регрессии.

Раздел 6. Обработка естественного языка.

Токенизация текстовых данных. Преобразование слов в их базовые формы с помощью стемминга. Преобразование слов в их корневые формы с помощью лемматизации. Разбиение текстовых данных на информационные блоки. Извлечение частотности слов с помощью модели Bag of Words.

Раздел 7. Вероятностный подход к обработке последовательных данных.

Последовательные данные. Обработка временных рядов с помощью библиотеки Pandas. Извлечение срезов временных данных. Операции над временными данными. Извлечение статистики из временных рядов. Генерация данных.

Раздел 8. Распознавание речи.

Работа со звуковыми сигналами. Визуализация аудиосигналов. Преобразование аудиосигналов. Генерирование аудиосигналов. Синтезирование звуков. Извлечение речевых признаков. Распознавание слов.

Раздел 9. Обнаружение и отслеживание объектов.

Библиотека OpenCV. Вычисление разности между кадрами. Отслеживание объектов с помощью цветных пространств. Отслеживание объектов путем вычитания фоновых изображений. Алгоритм CAMShift. Отслеживание объектов с использованием оптических потоков.

Раздел 10. Искусственные нейронные сети.

Создание и тренировка нейронной сети. Создание классификатора на основе перцептрона. Построение однослойной нейронной сети. Построение многослойной нейронной сети. Создание векторного квантизатора. Анализ последовательных данных с помощью рекуррентных нейронных сетей. Обучение с подкреплением. Создание окружения. Создание агента обучения. Сверточные нейронные сети. Архитектура CNN. Типы слоев CNN. Создание линейного регрессора на основе перцептрона.

5.2.2. Содержание дисциплины/модуля по видам учебных занятий.

Лекционные занятия

Таблица 5.2.1

№ п/п	Номер раздела дисциплины	Объем, час.			Тема лекции
		ОФО	ЗФО	ОЗФО	
1	1	1	-	-	Введение. Технологии анализа данных
2	2	1	-	-	Классификация и регрессия
3	3	1	-	-	Предсказательная аналитика на основе ансамблевого обучения

4	4	2	-	-	Распознавание образов
5	5	1	-	-	Генетические алгоритмы
6	6	2	-	-	Обработка естественного языка
7	7	1	-	-	Вероятностный подход к обработке последовательных данных
8	8	1	-	-	Распознавание речи
9	9	2	-	-	Обнаружение и отслеживание объектов
10	10	2	-	-	Искусственные нейронные сети
Итого:		14	-	-	

Практические занятия

Практические занятия учебным планом не предусмотрены.

Лабораторные работы

Таблица 5.2.2

№ п/п	Номер раздела дисциплины	Объем, час.			Тема лабораторной работы
		ОФО	ЗФО	ОЗФО	
1	1	2	-	-	Технологии анализа данных и машинное обучение
2	2	6	-	-	Предварительная обработка данных
3	2	4	-	-	Классификация и регрессия
4	3	6	-	-	Предсказательная аналитика на основе ансамблевого обучения
5	4	4	-	-	Распознавание образов с помощью обучения без учителя
6	4	6	-	-	Создание рекомендательных систем
7	5	2	-	-	Генетические алгоритмы
8	6	4	-	-	Обработка естественного языка
9	7	4	-	-	Обработка последовательных данных
10	8	4	-	-	Распознавание речи
11	9	2	-	-	Обнаружение и отслеживание объектов
12	10	6	-	-	Искусственные нейронные сети
Итого:		42	-	-	

Самостоятельная работа студента

Таблица 5.2.3

№ п/п	Номер раздела дисциплины	Объем, час.			Тема	Вид СРС
		ОФО	ЗФО	ОФО		
1	1	-	-	-	Введение. Технологии анализа данных	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
2	2	4	-	-	Классификация и регрессия	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
3	3	4	-	-	Предсказательная аналитика на основе ансамблевого обучения	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
4	4	8	-	-	Распознавание образов	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
5	5	4	-	-	Генетические алгоритмы	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
6	6	10	-	-	Обработка естественного языка	Подготовка к

						лабораторным работам, и оформление отчета по лабораторным работам
7	7	4	-	-	Вероятностный подход к обработке последовательных данных	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
8	8	6	-	-	Распознавание речи	Подготовка к лабораторным работам, подготовка к коллоквиуму
9	9	6	-	-	Обнаружение и отслеживание объектов	Подготовка к лабораторным работам, подготовка к коллоквиуму
10	10	6	-	-	Искусственные нейронные сети	Подготовка к лабораторным работам, и оформление отчета по лабораторным работам
11	1-10	36	-	-	Экзамен	Изучение вопросов и подготовка к экзамену
Итого:		88	-	-	X	X

5.2.3. Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

- визуализация учебного материала в PowerPoint в диалоговом режиме (лекционные занятия);
- работа в малых группах (лабораторные работы);
- индивидуальная работа (лабораторные работы).

6. Тематика курсовых работ

1. Оценка стоимости недвижимости с использованием регрессора на основе машины опорных векторов.
2. Прогнозирование интенсивности дорожного движения с помощью классификатора на основе предельно случайных лесов.
3. Сегментирование рынка на основе моделей совершения покупок.
4. Поиск пользователей с похожими предпочтениями методом коллаборативной фильтрации.
5. Анализ географических данных.
6. Создание решателя для прохождения лабиринта.
7. Создание контроллера интеллектуального робота.
8. Создание робота для игры «Крестики-нолики».
9. Создание двух роботов, играющих между собой в игру «Шесть пешек».
10. Тематическое моделирование с использованием латентного размещения Дирихле.
11. Поиск нечетких дубликатов текста.
12. Идентификация буквенных последовательностей с помощью случайных полей.
13. Анализ биржевого рынка.
14. Создание системы распознавания речи.
15. Преобразование текста в речь.
16. Обнаружение и отслеживание лиц.
17. Отслеживание глаз и определение координат взора.
18. Создание системы оптического распознавания символов.
19. Создание классификатора изображений на основе однослойной нейронной сети.

20. Создание классификатора изображений на основе сверточной нейронной сети.

7. Контрольные работы

Контрольные работы учебным планом не предусмотрены.

8. Оценка результатов освоения дисциплины

8.1. Критерии оценивания степени полноты и качества освоения компетенций в соответствии с планируемыми результатами обучения приведены в Приложении 1.

8.2. Рейтинговая система оценивания степени полноты и качества освоения компетенций обучающихся очной формы обучения представлена в таблице 8.1.

Таблица 8.1

№ п/п	Виды мероприятий в рамках текущего контроля	Количество баллов
1 текущая аттестация		
1	Лабораторная работа №1	0 – 5
2	Лабораторная работа №2	0 – 7
3	Лабораторная работа №3	0 – 8
ИТОГО за первую текущую аттестацию		0 – 20
2 текущая аттестация		
4	Лабораторная работа №4	0 – 7
5	Лабораторная работа №5	0 – 8
6	Лабораторная работа №6	0 – 10
7	Лабораторная работа №7	0 – 10
ИТОГО за вторую текущую аттестацию		0 – 35
3 текущая аттестация		
8	Лабораторная работа №8	0 – 15
9	Лабораторная работа №9	0 – 10
10	Коллоквиум	0 – 5
11	Лабораторная работа №10	0 – 15
ИТОГО за третью текущую аттестацию		0 – 45
ВСЕГО		100

9. Учебно-методическое и информационное обеспечение дисциплины

9.1. Перечень рекомендуемой литературы представлен в Приложении 2.

9.2. Современные профессиональные базы данных и информационные справочные системы:

- Электронный каталог/Электронная библиотека Тюменского индустриального университета <http://webirbis.tsogu.ru/>;
- Научно – техническая библиотека ФГАОУ ВО РГУ Нефти и газа (НИУ) им. И.М. Губкина и ФГБОУ ВО «ТИУ» <http://elib.gubkin.ru/>;
- Научно – техническая библиотека ФГБОУ ВО «УГНТУ» и ФГБОУ ВО «Тюменский индустриальный университет» <http://bibl.rusoil.net>;
- Научно – техническая библиотека ФГБОУ ВО «УГТУ» и ФГБОУ ВО «Тюменский индустриальный университет» <http://lib.ugtu.net/books>;

- База данных «Научная электронная библиотека eLIBRARY.RU» (эл.подписи);
- ООО «ЭБС ЛАНЬ» www.e.lanbook.ru;
- ООО «Издательство ЛАНЬ» www.e.lanbook.com;
- ООО «Электронное издательство ЮРАЙТ» www.urait.ru;
- База данных Консультант студента «Электронная библиотека технического ВУЗа», ООО «Политехресурс» <http://www.studentlibrary.ru>;
- ООО «КноРус медиа», <https://www.book.ru>;
- Электронно - библиотечная система «IPRbooks», ООО Компания «Ай Пи Ар Медиа» <http://www.iprbookshop.ru/>;
- Национальная электронная библиотека (через терминалы доступа).

9.3. Лицензионное и свободно распространяемое программное обеспечение, в т.ч. отечественного производства.

- Microsoft Windows;
- Microsoft Office Professional;
- Экспертная оболочка CLIPS (свободно-распространяемое ПО),
- Visual Studio Code (свободно-распространяемое ПО),
- Eclipse (свободно распространяемое ПО).

10. Материально-техническое обеспечение дисциплины/модуля

Помещения для проведения всех видов работы, предусмотренных учебным планом, укомплектованы необходимым оборудованием и техническими средствами обучения.

Таблица 10.1

№ п/п	Перечень оборудования, необходимого для освоения дисциплины	Перечень технических средств обучения, необходимых для освоения дисциплины (демонстрационное оборудование)
1	-	Комплект мультимедийного оборудования: проектор, экран, компьютер, акустическая система. Локальная и корпоративная сеть.

11. Методические указания по организации СРС

11.1. Методические указания по подготовке к лабораторным занятиям.

Лабораторные занятия способствуют углублённому изучению дисциплины и служат основной формой подведения итогов самостоятельной работы студентов. Основная цель лабораторных занятий заключается в том, чтобы не только углубить и закрепить теоретические знания, но и сформировать практические компетенции, необходимые будущим специалистам.

На лабораторных занятиях приветствуется активное участие в обсуждении конкретных ситуаций, способность на основе полученных знаний находить наиболее эффективные решения

поставленных проблем, уметь находить полезный дополнительный материал по тематике занятий.

Студенту рекомендуется следующая схема подготовки к занятию:

1. Проработать конспект лекций;
2. Изучить рекомендованную литературу;
3. При затруднениях сформулировать вопросы к преподавателю
4. После выполнения лабораторной работы оформит отчет и подготовиться к защите.

11.2. Методические указания по организации самостоятельной работы.

Самостоятельная работа является одной из важнейших форм изучения любой дисциплины. Она позволяет систематизировать и углубить теоретические знания, закрепить умения и навыки, способствует развитию умений пользоваться научной и учебно-методической литературой. Познавательная деятельность в процессе самостоятельной работы требует от студента высокого уровня активности и самоорганизованности.

В учебном процессе выделяют два вида самостоятельной работы: аудиторная и внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа студентов представляет собой логическое продолжение аудиторных занятий. Затраты времени на выполнение этой работы регламентируются рабочим учебным планом. Режим работы выбирает сам обучающийся в зависимости от своих способностей и конкретных условий.

Самостоятельная работа может осуществляться индивидуально или группами студентов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений студентов.

Самостоятельная работа включает в себя работу с конспектом лекций, изучение и конспектирование рекомендуемой литературы, подготовка мультимедиа-сообщений/докладов, подготовка реферата, тестирование, решение задач и упражнений по образцу, решение вариативных задач, выполнение чертежей, схем, расчетов (графических работ), решение ситуационных (профессиональных) задач, подготовка к деловым играм, проектирование и моделирование разных видов и компонентов профессиональной деятельности, научно-исследовательскую работу и др.

Контроль результатов внеаудиторной самостоятельной работы студентов может осуществляться в пределах времени, отведенного на обязательные учебные занятия по

дисциплине и внеаудиторную самостоятельную работу студентов по дисциплине, может проходить в письменной, устной или смешанной форме.

Планируемые результаты обучения для формирования компетенции и критерии их оценивания

Дисциплина: Анализ данных и машинное обучение
 Код, направление подготовки: 38.03.05 Бизнес – информатика
 Направленность: Информационные системы предприятия

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения				
		1 – 2	3	4	5	
ОПК – 4	Знать (З1) основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения	Не знает основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения	Знает на низком уровне основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения	Знает на среднем уровне основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения	Знает в совершенстве основные математические методы и алгоритмы, используемые при разработке и использовании методов анализа данных и машинного обучения	
	Уметь (У1) использовать математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения	Не умеет использовать математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения	Умеет на низком уровне использовать математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения	Умеет на среднем уровне использовать математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения	Умеет в совершенстве использовать математический аппарат при разработке и оценке эффективности методов и технологий анализа данных и машинного обучения	
	Владеть (В1) практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения	Не владеет практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения	Владеет на низком уровне практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения	Владеет на среднем уровне практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения	Знает в совершенстве практическими навыками реализации математических методов и алгоритмов при разработке алгоритмов и программ анализа данных и машинного обучения	

КАРТА обеспеченности дисциплины учебной и учебно-методической литературой

Дисциплина: Анализ данных и машинное обучение

Код, направление подготовки: 38.03.05 Бизнес - информатика

Направленность: Информационные системы предприятия

№ п/п	Название учебного, учебно-методического издания, автор, издательство, вид издания, год издания	Количество экземпляров в БИК	Контингент обучающихся, использующих указанную литературу	Обеспеченность обучающихся литературой, %	Наличие электронного варианта в ЭБС (+/-)
1	Воронова Л.И. Machine Learning: регрессионные методы интеллектуального анализа данных [Электронный ресурс]: учебное пособие/ Воронова Л.И., Воронов В.И. - Электрон. текстовые данные. - М.: Московский технический университет связи и информатики, 2018. - 82 с. http://www.iprbookshop.ru/81325.html .	ЭР*	30	100	+
2	Даг, Т. Релевантный поиск с использованием Elasticsearch и Solr / Т. Даг, Б. Джон. - [Б. м.] : ДМК Пресс, 2018. - 408 с. - URL: https://e.lanbook.com/book/111439 . - Режим доступа: для автор. пользователей. - ЭБС Лань	ЭР*	30	100	+
3	Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс]: учебное пособие/ Пальмов С.В. - Электрон. текстовые данные. - Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. - 127 с. - Режим доступа: http://www.iprbookshop.ru/75376.html .	ЭР*	30	100	+
4	Шелудько В.М. Язык программирования высокого уровня Python. Функции, структуры данных, дополнительные модули [Электронный ресурс]: учебное пособие/ Шелудько В.М. - Электрон. текстовые данные. - Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2017.- 107 с.- Режим доступа: http://www.iprbookshop.ru/87530.html	ЭР*	30	100	+

ЭР – электронный ресурс для автор. пользователей доступен через Электронный каталог/Электронную библиотеку ТИУ <http://webirbis.tsogu.ru/>

Заведующий кафедрой БИМ
« 27 » мая 2021г.

Директор БИК
« 27 » мая 2021г.
М.П.

