МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Волгоградский государственный технический университет»

Факультет электроники и вычислительной техники

УТВЕРЖДЕНО
Факультет электроники и вычислительной техники
Декан Авдеюк О.А.
г.

Машинное обучение и анализ данных

рабочая программа дисциплины (модуля, практики)

Закреплена за кафедрой	**Системы автоматизированного проектирования и поискового конструирования**
Учебный план	Направление 09.03.01 Информатика и вычислительная техника Профиль "Вычислительные машины, комплексы, системы и сети"
Профиль	
Квалификация	**Бакалавр**
Срок обучения	**4 года**

Форма обучения	**очная**	Общая трудоемкость	**4 ЗЕТ**
Виды контроля в семестрах:	экзамены 6 курсовые работы 6		

Семестр(Курс.Номер семестра на курсе)	**6(3.2)**		Итого	
	УП	ПП	УП	ПП
Лекции	16	16	16	16
Лабораторные	32	32	32	32
Итого ауд.	48	48	48	48
Контактная работа	48.35	48.35	48.35	48.35
Сам. работа	60	60	60	60
Часы на контроль	35.65	35.65	35.65	35.65
Практическая подготовка	0	0	0	0
Итого трудоемкость в часах	144	144	0	0

ЛИСТ ОДОБРЕНИЯ, СОГЛАСОВАНИЯ И АКТУАЛИЗАЦИИ РАБОЧЕЙ ПРОГРАММЫ

Разработчик(и) программы:

Заведующий кафедрой Щербаков М.В. д.т.н.

Рецензент(ы):
(при наличии)

Рабочая программа дисциплины (модуля, практики)

Машинное обучение и анализ данных

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника (приказ Минобрнауки России от 19.09.2017 г. № 929)

составлена на основании учебного плана:

Направление 09.03.01 Информатика и вычислительная техника
Профиль "Вычислительные машины, комплексы, системы и сети"

Профиль:

утвержденного учёным советом вуза от 05.06.2019 протокол № 12.

Рабочая программа одобрена на заседании кафедры

Системы автоматизированного проектирования и поискового конструирования

номер протокола 2019 г.
Зав. кафедрой Щербаков Максим Владимирович

СОГЛАСОВАНО:
Факультет электроники и вычислительной техники
Председатель НМС факультета: Авдеюк О.А.

Протокол заседания НМС от
г. №

Рабочая программа дисциплины (модуля, практики) актуализирована 31.08.2023

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ). ВИД, ТИП ПРАКТИКИ, СПОСОБ И ФОРМА (ФОРМЫ) ЕЕ ПРОВЕДЕНИЯ.
Цель изучения дисциплины - ознакомление студентов с теоретическими основами и основными принципами машинного обучения.
Задачи изучения дисциплины:
- изучение теоретических основ и основных принципов машинного обучения;
- формирование у студентов практических навыков работы с данными и решения прикладных задач анализа данных.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	
Цикл (раздел) ОП:	Б1.В
2.1	**Требования к предварительной подготовке обучающегося:**
2.1.1	Основы системного программного обеспечения
2.1.2	Введение в направление
2.2	**Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**
2.2.1	Выполнение и защита выпускной квалификационной работы
2.2.2	Производственная практика: Преддипломная практика
2.2.3	Киберправо
2.2.4	
2.2.5	
2.2.6	
2.2.7	
2.2.8	
2.2.9	

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)
ПК-5: Способен консультировать пользователей по работе с функциями системы, обрабатывать запросы пользователей на изменение требований к системе, разрабатывать эксплуатационные документы, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала
ПК-5.1: Знать: терминологическую лексику областей профессиональных знаний для консультирования пользователей по работе с функциями системы, методы обработки запросов пользователей на изменение требований к системе, основные понятия и приемы разработки эксплуатационных документов, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала.
Результаты обучения: знает подходы в рамках дисциплины "машинное обучение и анализ данных" к терминологической лексике областей профессиональных знаний для консультирования пользователей по работе с функциями системы, методам обработки запросов пользователей на изменение требований к системе, основным понятиям и приемам разработки эксплуатационных документов, адресованные конечному пользователю компьютерной системы, а также стандартным техническим документам на основе предоставленного материала.
ПК-5.2: Уметь: консультировать пользователей по работе с функциями системы, обрабатывать запросы пользователей на изменение требований к системе, разрабатывать эксплуатационные документы, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала.
Результаты обучения: умеет принимать теоретические и практические подходы в рамках дисциплины "машинное обучение и анализ данных" - консультировать пользователей по работе с функциями системы, обрабатывать запросы пользователей на изменение требований к системе, разрабатывать эксплуатационные документы, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала.
ПК-5.3: Владеть навыком: консультирования пользователей по работе с функциями системы, обработки запросов пользователей на изменение требований к системе, разработок эксплуатационных документов, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала.
Результаты обучения: владеет практическими навыками применения консультирования пользователей по работе с функциями системы, обработки запросов пользователей на изменение требований к системе, разработок эксплуатационных документов, адресованные конечному пользователю компьютерной системы, а также стандартные технические документы на основе предоставленного материала.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Форма контроля

1	**Раздел 1. Введение в задачи, методы и инструменты машинного обучения.**			
1.1	Введение в задачи, методы и инструменты машинного обучения. /Тема/	6	0	
1.1.1	Введение в задачи, методы и инструменты машинного обучения. /Лек/	6	2	Эк, КР
1.1.2	Базовые статистические инструменты анализа данных: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебных задач /Лаб/	6	4	Эк, КР
2	**Раздел 2. Разведочный анализ данных.**			
2.1	Разведочный анализ данных. /Тема/	6	0	
2.1.1	Протокол разведочного анализа данных. Распределения данных, статистические эксперименты и проверка значимости. /Лек/	6	2	Эк, КР
2.1.2	Разведочный анализ данных: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебных задач /Лаб/	6	4	Эк, КР
2.2	Регрессия в машинном обучении. /Тема/	6	0	
2.2.1	Простая и множественная линейная регрессия, нелинейная регрессия. /Лек/	6	2	Эк, КР
2.2.2	Регрессионный анализ: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.3	Введение в задачи классификации. /Тема/	6	0	
2.3.1	Наивный байесовский алгоритм, дискриминантный анализ и логистическая регрессия, /Лек/	6	2	Эк, КР
2.3.2	Метод главных компонент: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.4	Качество классификации /Тема/	6	0	
2.4.1	Оценивание моделей классификации и стратегии в отношении несбалансированных данных. Статистическое машинное обучение. /Лек/	6	2	Эк, КР
2.4.2	Деревья решений:Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.5	Информационно-центричные методы классификации /Тема/	6	0	
2.5.1	Метод k ближайших соседей, древовидные модели, бэггинг и случайный лес, бустинговый подход. /Лек/	6	2	Эк, КР
2.5.2	Ансамблевый подход: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.6	Искусственные нейронные сети /Тема/	6	0	
2.6.1	Область применения и основные виды ИНС, глубокое обучение /Лек/	6	2	Эк, КР
2.6.2	Искусственные нейронные сети: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.7	Обучение без учителя. /Тема/	6	0	
2.7.1	Анализ главных компонент, кластеризация на основе k средних и иерархическая. Модельно-ориентированная кластеризация /Лек/	6	2	Эк, КР
2.7.2	Кластеризация данных: Теоретический минимум, изучение базовых библиотечных инструментов R/Python, решение учебной задачи /Лаб/	6	4	Эк, КР
2.7.3	Анализ учебно-прикладных данных и решение задачи классификации /Ср/	6	30	Эк, КР
2.7.4	Анализ учебно-прикладных данных и решение задачи регрессии /Ср/	6	30	Эк, КР
2.7.5	Курсовая работа /КР/	6	5.65	Эк, КР
2.7.6	Консультация /КоРа/	6	0.35	Эк, КР
2.7.7	Подготовка к экзамену /Экзамен/	6	30	Эк, КР

Примечание. Формы контроля: Эк – экзамен, К- контрольная работа, Ко- контрольный опрос, Сз- семестровое задание, З-зачет, ОП-отчет по практике.

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства планируемых результатов обучения представлены в виде фондов оценочных средств (ФОС), разработанных в соответствии с локальным нормативным актом университета. ФОС может быть представлен в Приложении к рабочей программе.

Контрольные вопросы и задания
Базовые статистические инструменты анализа данных:

Протокол разведочного анализа данных.
Выявление выбросов
Проверка однородности групповых дисперсий
Проверка на нормальность распределений
Выявление коллинеарности и формы связи между переменными
Влияние пространственно-временных факторов на анализируемую переменную.
Прямоугольные данные
Оценки центрального положения и вариабельности
Распределение данных и анализ двоичных и категориальных данных
Корреляция
Случайный отбор и смещенная выборка
Выборочное распределение статистики
Бутстрап
Доверительные интервалы распределения
Основные теоретические распределения, их применимость и свойства
A/B тестирование
Проверка статистических гипотез и перестановочные тесты
Статистическая значимость и p-значения
Мощность и размер выборки
Простая линейная регрессия
Множественная линейная регрессия
Факторные переменные в регрессии
Нелинейная регрессия
Наивный байесовский алгоритм классификации
Дискриминантный анализ
Логистическая реграссия
Оценивание моделей классификации
Стратегии в случае несбалансированных данных
Метод к ближайших соседей
Древовидные модели
Бэггинг и случайный лес
Бустинг
Анализ главных компонент
Кластеризация на основе к средних
Иерархическая кластеризация
Модельно-ориентированная кластеризация
Шкалирование и категориальные переменные

Темы письменных работ (курсовая работа)

Курсовая работа по машинному обучению и анализу данных представляет собой законченную работу, включающую программные решения задач математико-статистического моделирования и машинного обучения моделей на языке R/Python. Выбор темы курсовой работы осуществляется в соответствии с перечнем заданий и утверждается преподавателем.

Курсовая работа состоит из 5 разделов/заданий по разделам.

1. изучение предметной области и постановка задачи
2. сбор и обработка данных
3. разведочный анализ данных
4. построение статистических моделей и алгоритмов их машинного обучения
5. интерпретация и использование модели (концепция продукта данных или принятие решения на основе обобщения данных)

В результате формируется пояснительная записка, содержащая результаты выполнения по разделам. Курсовая работа оформляется в соответствии с шаблоном и оценивается в процессе защиты работы. Оформляется задание на курсовую работу и пояснительная записка по шаблонам, приведённым на сайте ВолгГТУ: http://umu.vstu.ru/umu-docs/uo/forms
Каждый раздел оценивается максимум на 5 баллов.

Темы курсовых работ

1.Прогнозирование возврата персональных ссуд в инвестиционно-кредитной компании Lending Club (датафрейм loan200 из пакета FNN языка R)
2.Задача кредитного скоринга по данным о 1000 клиентов одного из немецких банков (датафрейм GermanCredit из пакета caret языка R)
3.Оценивание стоимости домов на основе данных о жилом округе Кинг (Сэтл, шт. Вашингтон) (датафрейм data.frame house в базовом оснащении языка R)
4.Задача об обнаружении спама по коллекции данных Центра машинного обучения и интеллектуальных систем Калифорнийского университета http://archive.ics.uci.edu/ml/
5.Использование результатов гидробиологических исследований обилия водорослей в различных реках (датафрейм algae из пакета DMwR языка R)
6. Изучение семейного статуса 2000 швейцарцев на протяжении 16 лет их жизни между 15 и 30 годами (датафрейм biofam из пакета TraMineR языка R)
7. Определение способа изготовления стекла по его химическому составу (датафрейм fgl из пакета MASS языка R)
8.Определение возраста морского ушка на данных о 4177 особях этого вида. Данные берутся из коллекции данных Центра

машинного обучения и интеллектуальных систем Калифорнийского университета http://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data
Геоботаника:
9.Анализ зависимости между показателями химического состава и обилием 44 видов травянистых растений на 24 пробных площадках (датафреймы varechem и varecpec из пакета vegan языка R)
10.Задача анализа и моделирования данных в различных предметных областях. На данных, найденных студентами и одобренных преподавателем.
Все темы подразумевают широкую вариабельность поставленных и решаемых задач. В процессе выполнения курсовой работы студентам необходимо продемонстрировать различные подходы к решению поставленных задач на основе широкого класса методов и моделей машинного обучения. Особенно высоко оценивается своеобразие постановки задачи и оригинальность ее решения, состоящее в уникальной последовательности операций преобразования и моделирования данных.

Фонды оценочных средств

Оценочные средства планируемых результатов обучения представлены в виде фондов оценочных средств (ФОС), разработанных в соответствии с локальным нормативным актом университета. ФОС представлен в Приложении к рабочей программе.

Перечень видов оценочных средств

Наименование оценочного средства 1: курсовая работа. Курсовая работа по машинному обучению и анализу данных представляет собой законченную работу, включающую программные решения задач математико-статистического моделирования и машинного обучения моделей на языке R/Python. Позволяет оценить умения обучающихся самостоятельно исследовать данные, решать задачи статистического моделирования и использовать построенные модели, судить о сформированности аналитических, исследовательских навыков, навыков практического и творческого мышления.

Наименование оценочного средства 2: Собеседование. 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. Собеседование применяется на экзамене.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)

6.1. Рекомендуемая литература

	Авторы, составители	Заглавие	Издательство, год.	Электронный адрес
Л.1	Афанасьева Т. В.	Введение в проектирование систем интеллектуального анализа данных: Учебное пособие	Ульяновский государственный технический университет, 2017	https://e.lanbook.com/book/165064
Л.2	Ахирвар К.	Состязательные сети. Проекты: учебное пособие	Москва: ДМК Пресс, 2020	https://e.lanbook.com/reader/book/140586/#7
Л.3	Паттерсон Дж., Гибсон А.	Глубокое обучение с точки зрения практика	Москва: ДМК Пресс, 2018	https://e.lanbook.com/reader/book/116122/#6
Л.4	Гудфеллоу Я., Бенджио И., Курвилль А.	Глубокое обучение	Москва: ДМК Пресс, 2018	https://e.lanbook.com/reader/book/107901/#1

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	https://ranalytics.github.io/data-mining/ - электронная книга Шитикова В.К., Мастицкого С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R.
Э2	ttps://r-analytics.blogspot.com/ - блог С. Мастицкого, посвященный вопросам, связанным со статистическим моделированием и машинным обучением
Э3	ttps://www.youtube.com/channel/UCLk-Oih8VlqF-StidijTUnw/featured - youtube-канал “Основы анализа данных”, содержащий курсы по анализу данных, статистике, эконометрике и R
Э4	https://www.youtube.com/channel/UCKFojzto0n4Ab3CRQRZ2zYA/featured - youtube-канал “Компьютерные науки”
Э5	ttps://www.youtube.com/channel/UCeq6ZIlvC9SVsfhfKnSvM9w/featured - youtube-канал “ML Trainings”, посвященный практике и соревнованиям по машинному обучению

6.3 Перечень программного обеспечения

6.3.1.1	Adobe Acrobat Reader DC — бесплатное решение для просмотра файлов PDF
6.3.1.2	LibreOffice — офисный пакет
6.3.1.3	The R Project for Statistical Computing - бесплатный язык программирования
6.3.1.4	RStudio - бесплатная среда разработки
6.3.1.5	Python - бесплатный язык программирования
6.3.1.6	Произвольная бесплатная IDE для Python

6.4 Перечень информационных справочных систем и электронных библиотечных систем (ЭБС)

6.3.2.1	Библиотека (НТБ), http://library.vstu.ru/sci-nci
6.3.2.2	Электронная информационно-образовательная среда университета,http://eos.vstu.ru

6.3.2.3	ЭБС "Лань", https://e.lanbook.com/
6.3.2.4	ЭБС "Book.ru", https://www.book.ru/
6.3.2.5	Электронная библиотека "Grebennikon", https://grebennikon.ru/

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ) /ОБОРУДОВАНИЕ

7.1	Мультимедийная учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. /Учебная доска, учебная мебель,интерактивная трибуна, видеопроектор.
7.2	Лаборатория информационных технологий. /Учебная мебель, компьютерная техника, оснащенная программным обеспечением, доступом в Интернет и в электронную информационно-образовательную среду университета.
7.3	Аудитория для самостоятельной работы обучающихся./Учебная мебель, компьютерная техника с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета (читальный зал информационно-библиотечного центра.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)

Организация образовательного процесса по данной дисциплине регламентируется учебным планом и расписанием учебных занятий. При формировании своей индивидуальной образовательной траектории обучающийся имеет право на перезачет дисциплины (переаттестации ее части), если она была освоена в процессе предшествующего обучения. Перезачёт (переаттестации ее части)освобождает обучающегося от необходимости повторного освоения дисциплины (полностью или частично).

Учебный процесс при преподавании курса основывается на использовании традиционных, инновационных и информационных образовательных технологий. Традиционные образовательные технологии представлены лекциями и практическими занятиями. Инновационные образовательные технологии используются в виде широкого применения активных и интерактивных форм проведения занятий. Информационные образовательные технологии реализуются путем активизации самостоятельной работы студентов в электронной информационной образовательной среде.

Лекционный курс предполагает систематизированное изложение основных вопросов учебного плана. На первой лекции лектор информирует студентов о рекомендуемой литературе и электронных источниках информации по дисциплине, с указанием, какой учебник (учебное пособие) является базовым.

Лабораторные работы предполагают выполнение и отчет заданий по темам, рассмотренным на лекционных занятиях. Каждому лабораторному занятию предшествует самостоятельная подготовка студента, включающая: ознакомление с содержанием лабораторной работы по методическим указаниям; проработку теоретической части по лекционному материалу и учебникам, рекомендованным в методических указаниях;

Самостоятельная работа студентов включает изучение законспектированного на лекционных занятиях материала, дополнение его с учетом рекомендованной по данной теме литературы, самостоятельную подготовку к лабораторным работам, самостоятельное выполнение и оформление заданий контрольной работы, аналогичных выполненным на занятиях.

Курсовой проект студентов включает разведочный анализ данных и моделирование на основе принципов машинного обучения (предметная область, задача и датасет определяются индивидуальным заданием)

Перечень методических указаний для освоения дисциплины представлен в таблице 6.1.3

В течении семестра для студентов проводятся групповые текущие консультации по учебной дисциплине, а также консультация перед экзаменом.

Методические рекомендации по обучению лиц с ограниченными возможностями здоровья и инвалидов

Профессорско-педагогический состав знакомится с психолого-физиологическими особенностями обучающихся инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ), индивидуальными программами реабилитации инвалидов (при наличии). При необходимости осуществляется дополнительная поддержка преподавания тьюторами, психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. N АК-44/05вн), в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе. Подбор и разработка учебных материалов производятся с учетом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем.

Освоение дисциплины лицами с ОВЗ осуществляется с использованием средств обучения общего и специального назначения (персонального и коллективного использования). Материально-техническое обеспечение предусматривает приспособление аудиторий к нуждам лиц с ОВЗ (при необходимости).