

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программы бакалавриата
по направлению подготовки
11.03.02 Инфокоммуникационные технологии и
системы связи,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Введение в машинное обучение и DataScience

Направление подготовки: 11.03.02 Инфокоммуникационные
технологии и системы связи

Направленность (профиль): Системы мобильной связи и сетевые
технологии на транспорте

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде
электронного документа выгружена из единой
корпоративной информационной системы управления
университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 170737
Подписал: заместитель директора академии Паринов Денис
Владимирович
Дата: 19.01.2024

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины «Введение в машинное обучение и DataScience» является формирование у обучающихся компетенций в соответствии с требованиями самостоятельно утвержденного образовательного стандарта высшего образования (СУОС) по направлению подготовки бакалавриата «Инфокоммуникационные технологии и системы связи».

Задачами дисциплины являются приобретение обучающимися знаний о способе построения информационных систем для решения неформализованных задач в различных сферах деятельности человека; умений использовать современные программные средства для моделирования знаний, принятия решений и навыков разработки экспертных систем.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ОПК-3 - Способен применять методы поиска, хранения, обработки, анализа и представления в требуемом формате информации из различных источников и баз данных, соблюдая при этом основные требования информационной безопасности.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

классификацию интеллектуальных информационных систем. Экспертные системы. Системы с интеллектуальным интерфейсом. Назначение и область применения естественно-языковых систем

Уметь:

осуществлять сбор и анализ исходных данных для формулирования задач разработки, расчета и проектирования систем и средств автоматизации, и управления; использовать навыки работы со структурой экспертных систем. Создавать и использовать интеллектуальные хранилища

Владеть:

навыками документирования и моделирования бизнес-процессов и технологических процессов объекта автоматизации; подходами к построению интеллектуальных информационных систем

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Сем. №5
Контактная работа при проведении учебных занятий (всего):	80	80
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа	64	64

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 64 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	Введение. Классификация интеллектуальных информационных систем. Цели преподавания дисциплины, терминология. Подходы к построению интеллектуальных информационных систем.
2	Характеристика основных направлений, по которым проводятся исследования. Характеристика основных направлений, по которым проводятся исследования. Особенности и признаки интеллектуальности
3	Экспертные системы История развития, особенности экспертных систем, цели их создания. Структура ЭС, классификация, назначение основных блоков. Этапы создания ЭС
4	Самообучающиеся системы. Системы с интеллектуальным интерфейсом Индуктивные систем. Системы на прецедентах. Интеллектуальные хранилища. Назначение и область применения естественно-языковых систем. Системы общения с базами данных

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Предварительная обработка массивов данных Массивы данных. Предварительная обработка массивов данных. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.
2	Алгоритмы классификации в анализе данных Описание и изучение алгоритмов классификации в анализе данных. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.
3	Алгоритмы регрессионного анализа данных Регрессионный анализ данных. Применение. Алгоритмы. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.
4	Алгоритмы кластерного анализа данных Кластерный анализ данных. Применение. Алгоритмы. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.
5	Разведочный анализ данных Представление наборов данных с помощью различных методов и библиотек, расчет значений в признаках, анализ и визуализация данных
6	Задача многоклассовой классификации сбалансированного набора данных Разбивка обработанного набора данных на обучающую и тестовую выборки, обучение моделей на основании библиотек, отображение точности работы каждой модели на тестовой и обучающей выборке
7	Классификация несбалансированного набора данных Произвести анализ данных и сделать выводы о структуре их набора. Произвести скалярную стандартизацию данных, разбить обработанный массив данных, произвести обучение моделей библиотеки. Отобразить точность работы каждой модели до и после кросс-валидации
8	Задача восстановления регрессии Произвести разведочный анализ с визуализацией данных. Произвести обучение моделей линейной и полиномиальной регрессии, отобразить точность каждой модели
9	Задача восстановления регрессии Отображение построенных моделей линейной и полиномиальной регрессии, выделив тестовую и обучающую выборки. Отобразить метрики моделей до и после кросс-валидации, произвести поиск оптимальных гиперпараметров с помощью методов и моделей с наилучшими показателями

№ п/п	Тематика практических занятий/краткое содержание
10	Задача кластеризации Произвести разведочный анализ данных с отображением, закодировать категориальный данных в дискретные величины, произвести скалярную стандартизацию данных
11	Задача кластеризации Произвести снижение размерности набора данных и обучение модели библиотеки по методу К-средних. Произвести разведочный анализ данных по кластерам для оценки качества обучения модели
12	Задача выявления аномалий Произвести разведочный анализ данных с отображением гистограммы распределения данных, определить особенности набора данных. Произвести кластеризацию данных и обучение моделей: метод К-средних, цепь Маркова, изолирующий лес, одноклассовый метод опорных векторов
13	Задача выявления аномалий Произвести разведочный анализ по набору данных для оценки качества обучения модели, определить причины аномалий. Провести дополнительный анализ данных, чтобы выявить причины аномалий и меры для их устранения
14	Задача прогнозирования временных рядов Произвести разведочный анализ с построением диаграммы распределения данных и диаграммы размаха, сделать выводы об особенностях данных, произвести кластеризацию данных и обучение модели XGBoost Regressor
15	Задача прогнозирования временных рядов Произвести тестовую и обучающую выборку данных для модели. Отобразить точность модели на тестовой и обучающей выборках. Сделать прогноз на очередной период на основании анализируемых данных модели, принять решение об улучшении параметров на основании прогноза

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Подготовка к практическим занятиям
2	Работа с лекционным материалом, литературой, самостоятельное изучение
3	Выполнение курсового проекта.
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых проектов

Курсовой проект на тему: «Синтез модели прогнозирования рисков задержки поездов из-за отказов объектов железнодорожной автоматики и телемеханики». Исходные данные выбираются в соответствии с вариантом:

Вариант №1

Величина: идентификатор; Тип величины: категориальная; Диапазон возможных значений: ID000-ID1000; Примечание: -.

Вариант №2

Величина: класс линии; Тип величины: категориальная; Диапазон возможных значений: Кодифируется числами 1 – 5; Примечание: Возможный признак группировки.

Вариант №3

Величина: специализация линии; Тип величины: категориальная; Диапазон возможных значений: Кодифируется числами 1 – 7; Примечание: Возможный признак группировки.

Вариант №4

Величина: количество элемент; Тип величины: количественная/целое число; Диапазон возможных значений: 1..N; Примечание: Фактор-причина.

Вариант №5

Величина: Количество пар поездов в сутки; Тип величины: Количественная / целое число; Диапазон возможных значений: 1...M; Примечание: Фактор-причина.

Вариант №6

Величина: Среднее время до восстановления; Тип величины: Количественная /дробное с округлением до десятых; Диапазон возможных значений: 0,1...Z; Примечание: Фактор-причина.

Вариант №7

Величина: Количество отступлений от норм содержания; Тип величины: Количественная / целое число; Диапазон возможных значений: 1...R; Примечание: Фактор-причина, данные содержат пропуски и возможные аномалии.

Вариант №8

Величина: Количество предотказных состояний; Тип величины: Количественная / целое число; Диапазон возможных значений: 1...X; Примечание: Фактор-причина, данные содержат пропуски и возможные аномалии.

Вариант №9

Величина: Количество отказов, вызывающих задержку в движении поездов; Тип величины: Количественная / целое число; Диапазон возможных значений: 1...F; Примечание: Фактор-следствие (одна из составляющих риска).

Вариант №10

Величина: Суммарные задержки поездов на один отказ; Тип величины: Количественная /дробное с округлением до десятых; Диапазон возможных

значений: 1...Р; Примечание: Фактор-следствие (вторая из составляющих риска).

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Python и анализ данных ISBN 978-5-97060-590-5 540 с. Маккинни У. Учебное пособие Издательство "ДМК Пресс" , 2020	https://e.lanbook.com/book/131721
2	Интеллектуальный анализ данных ISBN 978-5-94621-531-2 120 с. Замятин А.В. Учебное пособие Национальный исследовательский Томский государственный университет , 2016	https://e.lanbook.com/book/74565
3	Основы искусственного интеллекта ISBN 978-5-00101-908-4 130 с. Боровская Е. В., Давыдова Н. А. Учебное пособие Издательство "Лаборатория знаний" (ранее "БИНОМ. Лаборатория знаний") , 2020	https://e.lanbook.com/book/151502

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

Информационный портал Научная электронная библиотека eLIBRARY.RU (www.elibrary.ru);

Единая коллекция цифровых образовательных ресурсов (<http://window.edu.ru>);

Научно-техническая библиотека РУТ (МИИТ) (<http://library.miiit.ru>);

Поисковые системы «Яндекс», «Google» для доступа к тематическим информационным ресурсам;

Электронно-библиотечная система издательства «Лань» – <http://e.lanbook.com/>;

Электронно-библиотечная система ibooks.ru – <http://ibooks.ru/>;

Электронно-библиотечная система «УМЦ» – <http://www.umczdt.ru/>;

Электронно-библиотечная система «Intermedia» – <http://www.intermediapublishing.ru/>;

Электронно-библиотечная система «BOOK.ru» – <http://www.book.ru/>;

Электронно-библиотечная система «ZNANIUM.COM» – <http://www.znanium.com/>

Электронно-библиотечная система ibooks.ru – <http://ibooks.ru/> ;

Электронно-библиотечная система «УМЦ» – <http://www.umczdt.ru/>;

Электронно-библиотечная система «Intermedia» – [http:// www .intermediapublishing.ru/](http://www.intermediapublishing.ru/);

Электронно-библиотечная система «BOOK.ru» – <http://www.book.ru/>;

Электронно-библиотечная система «ZNANIUM.COM» – <http://www.znanium.com/>

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

Программное обеспечение для выполнения практических заданий включает в себя программные продукты общего применения: операционная система Windows, Microsoft Office 2003 и выше, Браузер Internet Explorer 8.0 и выше с установленным Adobe Flash Player версии 10.3 и выше, Adobe Acrobat, а также специализированное программное обеспечение Python, Anaconda, Jupiter NoteBook

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Рабочее место преподавателя с персональным компьютером, подключённым к сетям INTERNET

Специализированная лекционная аудитория с мультимедиа аппаратурой интерактивной доской

Компьютерный класс. Рабочие места студентов в компьютерном классе, подключённые к сетям INTERNET

Для проведения практических занятий: компьютерный класс; компьютеры с минимальными требованиями.

Технические требования к оборудованию для осуществления учебного процесса с использованием дистанционных образовательных технологий:

колонки, наушники или встроенный динамик (для участия в аудиоконференции); микрофон или гарнитура (для участия в аудиоконференции); веб-камеры (для участия в видеоконференции);

для ведущего: компьютер с процессором Intel Core 2 Duo от 2 ГГц (или аналог) и выше, от 2 Гб свободной оперативной памяти.

9. Форма промежуточной аттестации:

Зачет в 5 семестре.

Курсовой проект в 5 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

доцент, доцент, к.н. кафедры
«Системы управления транспортной
инфраструктурой»

И.А. Журавлев

Согласовано:

Заместитель директора академии

Д.В. Паринов

Председатель учебно-методической
комиссии

Д.В. Паринов