

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 Data Science

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: 167783
Подписал: руководитель образовательной программы
Киселёва Анастасия Сергеевна
Дата: 26.12.2024

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	Семестр №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. Рассматриваемые вопросы: - цели преподавания дисциплины, терминология; - подходы к построению интеллектуальных информационных систем.
2	Характеристика основных направлений, по которым проводятся исследования. Рассматриваемые вопросы: - характеристика основных направлений, по которым проводятся исследования; - особенности и признаки интеллектуальности.
3	Экспертные системы Рассматриваемые вопросы: - история развития, особенности экспертных систем, цели их создания; - структура ЭС, классификация, назначение основных блоков. Этапы создания ЭС.
4	Самообучающиеся системы. Системы с интеллектуальным интерфейсом Рассматриваемые вопросы: - индуктивные системы; - системы на прецедентах; - интеллектуальные хранилища; - назначение и область применения естественно-языковых систем; - системы общения с базами данных.
5	Кросс-валидация Рассматриваемые вопросы: - кросс-валидация.
6	Обработка категориальных данных Рассматриваемые вопросы: - применение методов кодирования (например, one-hot encoding) для работы с категориальными переменными.
7	Работа с текстовыми данными Рассматриваемые вопросы: - обработка и анализ текстовых данных с использованием методов NLP (например, TF-IDF, Bag of Words).
8	Оценка и выбор модели Рассматриваемые вопросы: - сравнение различных моделей машинного обучения и выбор наилучшей на основе метрик (точность, F1-score и др.).
9	Разработка проекта Data Science Рассматриваемые вопросы: - полный цикл проекта Data Science: от формулировки проблемы до развертывания модели и представления результатов.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Предварительная обработка массивов данных Рассматриваемые вопросы: - массивы данных. Предварительная обработка массивов данных. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.

№ п/п	Тематика практических занятий/краткое содержание
2	<p>Алгоритмы классификации в анализе данных Рассматриваемые вопросы: - описание и изучение алгоритмов классификации в анализе данных. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.</p>
3	<p>Алгоритмы регрессионного анализа данных Рассматриваемые вопросы: - регрессионный анализ данных. Применение. Алгоритмы. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.</p>
4	<p>Алгоритмы кластерного анализа данных Рассматриваемые вопросы: - кластерный анализ данных. Применение. Алгоритмы. Индивидуальное задание выбирается согласно методическим указаниям к семинарским занятиям.</p>
5	<p>Разведочный анализ данных Рассматриваемые вопросы: - представление наборов данных с помощью различных методов и библиотек, расчет значений в признаках, анализ и визуализация данных</p>
6	<p>Задача многоклассовой классификации сбалансированного набора данных Рассматриваемые вопросы: - разбивка обработанного набора данных на обучающую и тестовую выборки; - обучение моделей на основании библиотек; - отображение точности работы каждой модели на тестовой и обучающей выборке.</p>
7	<p>Классификация несбалансированного набора данных Рассматриваемые вопросы: - 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.</p>
8	<p>Задача восстановления регрессии Рассматриваемые вопросы: - произвести разведочный анализ с визуализацией данных; - произвести обучение моделей линейной и полиномиальной регрессии; - отобразить точность каждой модели.</p>
9	<p>Задача восстановления регрессии Рассматриваемые вопросы: - 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.</p>
10	<p>Задача кластеризации Рассматриваемые вопросы: - произвести разведочный анализ данных с отображением; - закодировать категориальный данных в дискретные величины; - произвести скалярную стандартизацию данных.</p>
11	<p>Задача кластеризации Рассматриваемые вопросы: - произвести снижение размерности набора данных и обучение модели библиотеки по методу K-средних; - произвести разведочный анализ данных по кластерам для оценки качества обучения модели.</p>

№ п/п	Тематика практических занятий/краткое содержание
12	<p>Задача выявления аномалий</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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	<p>Задача выявления аномалий</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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	<p>Задача прогнозирования временных рядов</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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	<p>Задача прогнозирования временных рядов</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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	Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5.	https://e.lanbook.com/book/131721
2	Платонова, О. В. Алгоритмические основы обработки данных : учебное пособие / О. В. Платонова, Ю. С. Асадова, А. А. Рыжова. — Москва : РТУ МИРЭА, 2024. — 101 с. — ISBN 978-5-7339-2174-7.	https://e.lanbook.com/book/421061
3	Бураков, Д. П. Основы хранения данных : учебное пособие / Д. П. Бураков. — Санкт-Петербург : ПГУПС, 2022. — 60 с. — ISBN 978-5-7641-1731-7.	https://e.lanbook.com/book/264668

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.miit.ru>);

Поисковые системы «Яндекс» 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/>

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).

1. Рабочее место преподавателя с персональным компьютером, подключённым к сетям INTERNET

2. Специализированная лекционная аудитория с мультимедиа аппаратурой интерактивной доской

3. 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:

Директор

Б.В. Игольников

Руководитель образовательной
программы

А.С. Киселёва

Председатель учебно-методической
комиссии

Д.В. Паринов