

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 магистратуры
по направлению подготовки
09.04.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Машинное обучение и анализ данных

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Технологии проектирования программного обеспечения

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: 5665
Подписал: заведующий кафедрой Нутович Вероника
Евгеньевна
Дата: 08.06.2022

1. Общие сведения о дисциплине (модуле).

Цель дисциплины «Машинное обучение и анализ данных» продемонстрировать студенту процесс создания интеллектуальной модели машинного обучения. Задачи данной дисциплины:

- освоение принципов сбора данных;
- освоения методов очистки и подготовки данных;
- освоение методов и инструментов разметки данных;
- освоение инструментов развертывания данных;
- освоение принципов подбора ключевых признаков;
- освоение принципов подбора алгоритмов машинного обучения;
- освоение инструментов обучения модели искусственного интеллекта;
- освоение методов оценки модели искусственного интеллекта;
- освоение подходов развертывания модели искусственного интеллекта.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-1 - Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте;

ОПК-2 - Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач;

ПК-3 - Способен использовать существующие методы и алгоритмы решения задач машинного обучения и обработки больших данных;

ПК-13 - Способность к решению актуальных научных задач, к получению новых научных результатов.

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

Знать:

- принципы и методы машинного обучения и анализа данных;
- основы разметки данных;
- алгоритмы машинного обучения;
- методы оценки, мониторинга и визуализации модели искусственного интеллекта;

- принципы развертывания модели искусственного интеллекта.

Уметь:

- очищать и подготавливать данные;
- размечать данные;
- разворачивать данные;
- определять ключивые признаки;
- обучать модель искусственного интеллекта;
- оценивать модель искусственного интеллекта;
- разворачивать модель искусственного интеллекта.

Владеть:

- навыком разработки модели искусственного интеллекта на основе витрины данных для дальнейшего развертывания.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 14 з.е. (504 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		
		№1	№2	№3
Контактная работа при проведении учебных занятий (всего):	104	36	32	36
В том числе:				
Занятия лекционного типа	52	18	16	18
Занятия семинарского типа	52	18	16	18

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 400 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	Сбор данных. Рассматриваемые вопросы: - концепции ETL/ELT; - основы работы с различными форматами данных; - основы работы с различными источниками данных; - принципы построения конвейера данных; - инструменты сбора данных.
2	Очистка и подготовка данных. Рассматриваемые вопросы: - основы процесса очистки данных; - основы процесса насыщения данных; - основы распределенных хранилищ данных; - инструменты очистки и подготовки данных.
3	Разметка данных. Рассматриваемые вопросы: - основы работы с различными медиа-форматами данных; - основы разметки текстовых данных; - основы разметки изображений; - основы разметки аудио-данных; - основы разметки видео-данных; - инструменты разметки данных.
4	Развертывание данных. Рассматриваемые вопросы: - основы работы с распределенными хранилищами данных; - основы работы с базами данных; - инструменты развертывания данных.
5	Определение ключивых признаков. Рассматриваемые вопросы: - понятие признака; - моделирование признаков; - методики выделения и отбора признаков; - проблемы переобучения.
6	Подбор алгоритма искусственного интеллекта. Рассматриваемые вопросы: - понятие алгоритма;

№ п/п	Тематика лекционных занятий / краткое содержание
	- алгоритмы различных видов аналитики; - различные модели нейронных сетей.
7	Обучение модели искусственного интеллекта. Рассматриваемые вопросы: - библиотеки и фреймворки для машинного обучения: Numpy, Pandas, Scipy, Scikit-learn, Theano, TensorFlow, Keras, PyTorch; - основы работы с Jupyter Notebook.
8	Оценка модели искусственного интеллекта. Рассматриваемые вопросы: - методики оценки модели искусственного интеллекта; - основы инструментов мониторинга работы модели искусственного интеллекта; - основы инструментов оценки работы модели искусственного интеллекта; - основы визуализации работы модели искусственного интеллекта.
9	Развертывание модели искусственного интеллекта. Рассматриваемые вопросы: - основные форматы моделей искусственного интеллекта; - основы работы с хранилищами моделей искусственного интеллекта; - принципы MLOps; - основы работы платформами MLOps.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Сбор данных. В рамках выполнения лабораторных работ студент получит навыки сбора данных в соответствии с концепцией ETL/ELT.
2	Сбор данных. В рамках выполнения лабораторных работ студент получит навыки построения конвейера данных.
3	Очистка и подготовка данных. В рамках выполнения лабораторных работ студент получит навыки работы с распределенными хранилищами данных.
4	Очистка и подготовка данных. В рамках выполнения лабораторных работ студент получит навыки работы с инструментами очистки данных.
5	Очистка и подготовка данных. В рамках выполнения лабораторных работ студент получит навыки организации корпоративных хранилищ данных.
6	Развертывание данных. В рамках выполнения лабораторных работ студент получит навыки развертывания данных с помощью Hadoop и Apache Hive.
7	Разметка данных. В рамках выполнения лабораторных работ студент получит навыки разметки табличных данных.
8	Разметка данных. В рамках выполнения лабораторных работ студент получит навыки разметки табличных данных.
9	Определение ключевых признаков. В рамках выполнения лабораторных работ студент получит навыки выделения и отбора первичных и

№ п/п	Наименование лабораторных работ / краткое содержание
	вторичных признаков.
10	Определение ключевых признаков. В рамках выполнения лабораторных работ студент получит навыки упрощения потенциальной модели с помощью снижения размерности.
11	Подбор алгоритма искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки применения алгоритмов различных видов аналитики для решений задач.
12	Подбор алгоритма искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки применения различных видов нейронных сетей для решений задач.
13	Подбор алгоритма искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки применения ансамблей алгоритмов для решений задач.
14	Обучение модели искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки работы с библиотеками и фреймворками машинного обучения.
15	Обучение модели искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки работы с Jupyter Notebook и витринами данных.
16	Оценка модели искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки работы с инструментами мониторинга, оценки и визуализации работы модели искусственного интеллекта.
17	Развертывание модели искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки экспорта обученной модели искусственного интеллекта в различных форматах в хранилище моделей.
18	Развертывание модели искусственного интеллекта. В рамках выполнения лабораторных работ студент получит навыки работы с платформами MLOps.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Изучение рекомендованной литературы.
2	Подготовка к лабораторным занятиям.
3	Выполнение курсового проекта.
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

4.4. Примерный перечень тем курсовых проектов

1. Разработка модели искусственного интеллекта по распознаванию автомобилей на изображении.
2. Разработка модели искусственного интеллекта по распознаванию дорожных знаков на изображении.
3. Разработка модели искусственного интеллекта по распознаванию

животных на изображении.

4. Разработка модели искусственного интеллекта по распознаванию поездов на изображении.

5. Разработка модели искусственного интеллекта по распознаванию самолетов на изображении.

6. Разработка модели искусственного интеллекта по прогнозированию цены продукта на основе табличных данных.

7. Разработка модели искусственного интеллекта по прогнозированию качества продукта на основе табличных данных.

8. Разработка модели искусственного интеллекта по прогнозированию удовлетворенности пассажиров авиакомпании на основе табличных данных.

9. Разработка модели искусственного интеллекта по прогнозированию курса иностранной валюты на основе табличных данных.

10. Разработка модели искусственного интеллекта по прогнозированию автомобильных пробок на основе табличных данных.

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

№ п/п	Библиографическое описание	Место доступа
1	У. Маккинни Python и анализ данных. Москва : ДМК Пресс, 2020. - 540 с. - ISBN 978-5-97060-590-5.	https://e.lanbook.com/book/131721
2	А. Г. Храмов Методы и алгоритмы интеллектуального анализа данных : учебное пособие. Самара : Самарский университет, 2019. - 176 с. - ISBN 978-5-7883-1414-3. Учебное пособие	https://e.lanbook.com/book/148603
3	Л. А. Демидова Интеллектуальный анализ данных на языке Python : учебно-методическое пособие. Москва : РТУ МИРЭА, 2021. - 92 с. Учебно-методическое издание	https://e.lanbook.com/book/218693
4	О. Н. Сапрыкин Интеллектуальный анализ данных : учебное пособие. Самара : Самарский университет, 2020. - 80 с. - ISBN 978-5-7883-1563-8. Учебное пособие	https://e.lanbook.com/book/188906
5	А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь Большие данные. Big Data. Санкт-Петербург : Лань, 2022. - 188 с. - ISBN 978-5-8114-9690-7.	https://e.lanbook.com/book/198599
6	М. М. Железнов Методы и технологии обработки больших данных : учебно-методическое пособие. Москва : МИСИ – МГСУ, 2020. - 46 с. - ISBN 978-5-	https://e.lanbook.com/book/145102

	7264-2193-3. Учебно-методическое издание	
7	Н. А. Бутаков, М. В. Петров, Д. Насонов Обработка больших данных с Apache Spark : учебно-методическое пособие. Санкт-Петербург : НИУ ИТМО, 2019. - 50 с. Учебно-методическое издание	https://e.lanbook.com/book/136573
8	А. С. Лебедев, Ш. Г. Магомедов Методы Big Data : учебно-методическое пособие. Москва : РТУ МИРЭА, 2021. - 91 с. Учебно-методическое издание	https://e.lanbook.com/book/182452
9	А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь Системы поддержки принятия решений : учебное пособие для вузов. Санкт-Петербург : Лань, 2021. - 108 с. - ISBN 978-5-8114-8489-8. Учебное пособие	https://e.lanbook.com/book/176903
10	Э. Гласснер Глубокое обучение без математики : справочник. Москва : ДМК Пресс, 2019. - 578 с. - ISBN 978-5-97060-701-5. Справочное пособие	https://e.lanbook.com/book/131696
11	Ю. Лю Обучение с подкреплением на PyTorch. Сборник рецептов : руководство. Москва : ДМК Пресс, 2020. - 282 с. - ISBN 978-5-97060-853-1.	https://e.lanbook.com/book/179493

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

Электронно-библиотечная система Научно-технической библиотеки МИИТ (<http://library.miit.ru/>)

Курсы Microsoft (<https://www.microsoft.com/ru-ru/learning/windows-training.aspx>)

Документация по Apache Camel (<https://camel.apache.org/docs/>)

Документация по Apache Nifi (<https://nifi.apache.org/docs.html>)

Документация по Apache Kafka (<https://kafka.apache.org/documentation/>)

Документация по Apache Airflow (<https://airflow.apache.org/docs/>)

Документация по Apache Hadoop (<https://hadoop.apache.org/docs/stable/>)

Документация по OpenRefine (<https://openrefine.org/documentation.html>)

Документация по Label Studio (<https://labelstud.io/guide/>)

Список инструментов по разметке данных (<https://github.com/heartexlabs/awesome-data-labeling>)

Документация по MLFlow (<https://mlflow.org/docs/latest/index.html>)

Документация по Grafana (<https://grafana.com/docs/>)

Документация по PostgresPro (<https://www.postgrespro.ru/docs>)

Документация по Jupyter Notebook (<https://docs.jupyter.org/en/latest/>)

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

Прикладное программное обеспечение

Microsoft Office

Apache Camel

Apache Nifi

Apache Kafka

Apache Airflow

PostgresPro

OpenRefine

Hadoop

awesome-data-labelling

Label-studio

Apache Hive

Python 3.10

PyCharm

Jupyter Notebook

MLflow

Grafana

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

Для лекционных занятий – наличие проектора и экрана.

Для лабораторных занятий – наличие персональных компьютеров вычислительного класса, сервер с графическим ускорителем с ядрами CUDA.

9. Форма промежуточной аттестации:

Зачет в 1, 2 семестрах.

Курсовой проект во 2 семестре.

Экзамен в 3 семестре.

10. Оценочные материалы.

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

Авторы:

старший преподаватель кафедры
«Цифровые технологии управления
транспортными процессами»

И.С. Разживайкин

Согласовано:

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

Н.А.Клычева