

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

АННОТАЦИЯ К
РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Анализ данных

Направление подготовки: 09.03.02 – Информационные системы и технологии

Направленность (профиль): Технологии искусственного интеллекта в транспортных системах

Форма обучения: Очная

Общие сведения о дисциплине (модуле).

Дисциплина «Анализ данных» формирует критически важный инженерный компетенс в конвейере машинного обучения – способность превращать сырые массивы телеметрической и инфраструктурной информации в качественные математические объекты. В условиях жесткого импортозамещения и цифровизации транспортной отрасли РФ рынок испытывает острый дефицит специалистов, способных проводить глубокий разведочный анализ, устранять аномалии и конструировать информативные признаки без использования проприетарных облачных AutoML-платформ. Студент погружается в реальный производственный процесс предиктивной аналитики, работая с зашумленными временными рядами и полуструктурированными логами. На практике осваивается полный цикл прикладного анализа данных – от первичного профилирования и статистической проверки гипотез до генерации признаков и оформления инженерного паспорта данных. Выпускник получает подтвержденный портфель воспроизводимых решений, что делает его конкурентоспособным

кандидатом на позиции аналитика данных в ведущих транспортных и ИТ-компаниях.

Целью освоения дисциплины является формирование у обучающихся системных теоретических знаний и прикладных инженерных умений в области разведочного анализа данных и конструирования признаков для обеспечения высокого качества входной информации, необходимой для обучения моделей искусственного интеллекта в транспортных системах.

Для достижения поставленной цели в рамках дисциплины решается комплекс задач, направленных на формирование у обучающихся способности: осуществлять комплексное профилирование и аудит качества многомерных массивов данных, применять статистически обоснованные стратегии импутации пропусков и фильтрации аномалий, конструировать многомерные визуализации для выявления скрытых закономерностей, выполнять корреляционный анализ и проверку статистических гипотез, генерировать производные признаки с жестким соблюдением требования отсутствия утечек данных, а также оформлять результаты анализа в виде воспроизводимого интерактивного отчета и технического паспорта данных.

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