

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ**

**ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

**(ДГТУ)**

Кафедра «Организация перевозок и дорожного движения»

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ**

**ДЛЯ ЛАБОРАТОРНЫХ РАБОТ ПО ДИСЦИПЛИНЕ**

**«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ НА ТРАНСПОРТЕ»**

Направление подготовки 23.04.01 «Технология транспортных процессов»

Программа: «Транспортная логистика», «Организация перевозок на автомобильном транспорте», «Интеллектуальные транспортные системы»

Ростов-на-Дону

ДГТУ

2023

УДК 656

Составитель:

Методические указания для лабораторных работ по дисциплине «Интеллектуальный анализ данных на транспорте» –

Ростов-на-Дону: Донской гос. техн. ун-т, 2023. – 16 с.

УДК 656

Методические указания предназначены для магистрантов очной и заочной формы обучения по направлению подготовки 23.04.01 «Технология транспортных процессов», программы «Транспортная логистика», «Организация перевозок на автомобильном транспорте», «Интеллектуальные транспортные системы».

Печатается по решению редакционно-издательского совета Донского государственного технического университета

Ответственный за выпуск зав. кафедрой «Организации перевозок и дорожного движения» д-р техн. наук, профессор В.В. Зырянов

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

В печать \_\_\_.\_\_\_.2023 г.

Формат 60×84/16. Объем \_\_\_ усл. п. л.

Тираж \_\_\_ экз. Заказ № \_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Издательский центр ДГТУ

Адрес университета и полиграфического предприятия:

344000, г. Ростов-на-Дону, пл. Гагарина, 1

© Донской государственный

технический университет, 2023

**Введение**

Интеллектуальный анализ данных как одно из новых междисциплинарных направлений включает в себя различные способы работы с данными и в зависимости от предметной области имеет некоторую специфику применения этих способов.

Ставшее уже классическим data mining применимо не ко всем задачам технологии транспортных процессов. В частности, исследование транспортных потоков требует специфических подходов к очистке, агрегированию, преобразованию и последующему анализу данных о потоке.

Знания, полученные при изучении «Математических методов в транспортных исследованиях», в полной мере могут быть применены при интеллектуальном анализе данных о транспортных потоках.

Последовательное выполнение лабораторных работ по дисциплине «Интеллектуальный анализ данных на транспорте» позволит получить необходимые магистранту навыки научных исследований и компетенции.

Лабораторные работы проводятся с применением информационных технологий для индивидуального набора данных (dataset). Результат выполнения лабораторных работ оформляется в виде отчета в соответствии с ГОСТ 2.105-95. Все вычисления и графики выполняются в электронных таблицах (например, MS Exсel), текст отчета оформляется с помощью текстовых редакторов (например, MS Word) в соответствии с примерами из методических указаний для практических занятий. Для защиты лабораторных работ представляется два файла (текст отчета в формате .docx и необходимые расчеты в формате .xlsx).. В имени файлов должны содержаться фамилия студента, группа и номер зачетки. Например: Иванов\_АМТЛ21\_123456789.docx . Отчет должен быть распечатан и сдан преподавателю.

Для выполнения лабораторных работ следует получать исходные данные у преподавателя

**Лабораторная работа №1**

**Подготовка и предварительный анализ транспортных данных**

**Цель работы**: получить навык подготовки к анализу и проведения предварительного статистического анализа транспортных данных.

**Порядок выполнения работы**

1. Получить файл исходных данных. Привести его к табличному виду. Провести идентификацию записей.
2. Провести очистку и преобразование данных для статистического анализа данных с обязательным заполнением ведомости очистки и преобразований.
3. Провести статистическое оценивание, описательный, дисперсионный и корреляционный анализ данных.

**Содержание отчета:**

1. Постановка задачи. Характеристика исходного набора данных.
2. Ведомость очистки и преобразований.
3. Обоснование выбора данных для статистического оценивания. Гистограммы, эмпирические функции распределения, параметры вероятностных распределений, проверка согласия эмпирических и теоретических распределений. Выводы из полученных результатов.
4. Таблицы с описательными статистиками для каждого параметра. Выводы из полученных результатов.
5. Обоснование выбора данных для дисперсионного анализа. Применение критерия Фишера. Выводы из полученных результатов.
6. Обоснование выбора параметров для корреляционного анализа. Корреляционные матрицы. Проверка значимости коэффициентов корреляции. Выводы из полученных результатов.
7. Вывод и ответы на контрольные вопросы.

**Контрольные вопросы.**

1. Какими средствами проводилась очистка данных?
2. Какие надстройки позволяют получить описательные статистики?
3. Какие распределения используются для проверки статистических гипотез?
4. Как очистка данных может повлиять на результат дисперсионного и корреляционного анализа?

**Лабораторная работа №2**

**Сравнение регрессионной и гармонической модели при описании транспортных данных**

**Цель работы**: получить навык сравнения двух видов моделей: регрессионных и гармонических применительно к транспортным данным, зависящим от времени

**Порядок выполнения работы**

1. Подготовить набор данных для проведения работы.
2. Определить факторы и функции отклика для построения моделей.
3. Построить регрессионную модель и провести ее анализ.
4. Построить гармоническую модель, провести ее анализ.
5. Сравнить регрессионную и гармоническую модель по критерию Фишера.

**Содержание отчета:**

1. Постановка задачи. Характеристика исходного набора данных.
2. Обоснование выбора факторов и откликов.
3. Результаты построения и анализа регрессионной модели.
4. Результаты построения и анализа гармонической модели.
5. Результат сравнения двух моделей.
6. Вывод и ответы на контрольные вопросы.

**Контрольные вопросы.**

1. Какие преобразования были проведены для подготовки данных к анализу?
2. Какие надстройки позволяют получить регрессионную модель?
3. Какие распределения использованы для проверки статистических гипотез?
4. Как выбор факторов и откликов может повлиять на результат данной лабораторной работы?

**Лабораторная работа №3**

**Применение методов глубокого обучения к транспортным данным**

**Цель работы**: получить навык применения методов глубокого обучения и их комбинаций к транспортным данным.

**Порядок выполнения работы**

1. Провести подготовку данных для анализа.
2. Провести предварительный анализ данных, преобразования шкалы измерений при необходимости.
3. Провести кластеризацию и классификацию с помощью методов k-means, k-medians, kNN, NB и их композиций. Параметры выбрать самостоятельно.

**Содержание отчета:**

1. Постановка задачи. Характеристика исходного набора данных.
2. Ведомость очистки и преобразований.
3. Обоснование выбора данных применения методов глубокого обучения.
4. Результаты применения методов k-means, k-medians, kNN, NB и их композиций. Выводы из полученных результатов.
5. Вывод и ответы на контрольные вопросы.

**Контрольные вопросы.**

1. Какими средствами проводилась очистка данных?
2. Какие достоинства и недостатки методов выявлены в результате применения к транспортным данным?
3. Как шкала измерений влияет на результат классификации по методу NB?