

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:  
Директор  
А. В. Замятин

Оценочные материалы по дисциплине

Введение в интеллектуальный анализ данных

по направлению подготовки

**10.03.01 Информационная безопасность**

Направленность (профиль) подготовки:  
**Безопасность компьютерных систем**

Форма обучения  
**Очная**

Квалификация  
**Бакалавр**

Год приема  
**2025**

СОГЛАСОВАНО:  
Руководитель ОП  
В.Н. Тренькаев

Председатель УМК  
С.П. Сущенко

Томск – 2024

## **1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами**

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1. Способен оценивать роль информации, информационных технологий и информационной безопасности в современном обществе, их значение для обеспечения объективных потребностей личности, общества и государства.

ОПК-1.3. Способен обеспечивать защиту информации при работе с базами данных, при передаче по компьютерным сетям.

ОПК-2. Способен применять информационно-коммуникационные технологии, программные средства системного и прикладного назначения, в том числе отечественного производства, для решения задач профессиональной деятельности.

ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1. Учитывает современные тенденции развития информационных технологий в своей профессиональной деятельности.

ИОПК-1.3.1. Понимает модели и структуры данных, физические модели баз данных, принципы организации и методы проектирования баз данных, языки и системы программирования баз данных.

ИОПК-2.4. Формулирует предложения по применению современных информационно-коммуникационных технологий, используемых для решения задач профессиональной деятельности.

ИОПК-3.2. Осуществляет применение основных понятий, фактов, концепций, принципов математики и информатики для решения задач профессиональной деятельности.

ИОПК-3.3. Выявляет научную сущность проблем, возникающих в ходе профессиональной деятельности, и применяет соответствующий математический аппарат для их формализации, анализа и выработки решения.

## **2. Оценочные материалы текущего контроля и критерии оценивания**

Элементы текущего контроля:

- реферат;
- лабораторные работы.

Примеры тем рефератов (ИОПК-1.1, ИОПК-1.3.1, ИОПК-2.4, ИОПК-3.2, ИОПК-3.3):

1. Современные нейронные сети в обработке данных (изображений, видео, технологических сигналов, музыки и т.п.)
2. Современные алгоритмы классификации (изображений, текстов и т.п.)
3. Интеллектуальная обработка данных в ... (промышленности, медицине, бизнесе, индустрии развлечений, досуга и др.)
4. Извлечение знаний из текстов
5. Детектирование аномалий
6. Разновидности сверточных нейронных сетей (на примере одной архитектуры: BERT)
7. Интеллектуальные алгоритмы в ранней диагностике заболеваний
8. Интеллектуальные алгоритмы в персонализированной медицине
9. Интеллектуальные алгоритмы в робототехнике, транспортных системах и т.п.
10. Интеллектуальные алгоритмы в банковском деле/страховании/...

Критерии оценивания:

Результаты подготовки и защиты реферата определяются оценками «зачтено», «не зачтено».

Оценка «зачтено» выставляется, если реферат подготовлен на достаточно высоком методическом уровне, а тема в достаточной степени раскрыта в пояснительной записке.

Примеры заданий лабораторных работ (ИОПК-1.1, ИОПК-1.3.1, ИОПК-3.2, ИОПК-3.3):

**Задание №1: Проведение разведочного анализа данных.**

**Цель:** Провести разведочный анализ выбранного набора данных.

**Задачи:** Выбрать набор табличных данных (например на сайте kaggle.com), набор данных должен содержать не менее 10 столбцов и 1000 строк различных типов. Определить характеристики набора данных (типы данных, размер набора данных, наличие пропусков, дубликатов, категориальных переменных). Провести корреляционный анализ признаков и вывести корреляционную матрицу. Построить различные визуализации для признаков (не менее 3 различных). Сделать выводы по полученным результатам.

**Описание отчета:** Название «№Лабораторной\_Ф\_И\_И\_№группы», отчет в формате html/pdf должен содержать исходный код, полученные результаты, комментарии, выводы.

**Задание №2: Предварительная обработка данных и оптимизация признакового пространства.**

**Цель:** провести предварительную обработку и оптимизировать признаковое пространство.

**Задачи:** выбрать набор табличных данных (например на сайте Kaggle.com), набор данных должен содержать пропуски и категориальные переменные. Оценить наличие пропусков, дубликатов, категориальных переменных и их характеристики. Построить корреляционную матрицу, оценить возможность использовать ее для отбора признаков. Заполнить пропуски в данных (попробовать 3 различных способа). Закодировать категориальные переменные (попробовать 3 различных способа). Оценить наличие выбросов (z-оценка или ящик с усами), удалить выбросы. Использую метод главных компонент, сжать признаковое пространство (до 3х компонент), визуализировать результат относительно целевой переменной (если целевой переменной нет, то в качестве ее взять один из категориальных признаков). Сделать выводы по полученным результатам.

**Описание отчета:** Название «№Лабораторной\_Ф\_И\_И\_№группы», отчет в формате html/pdf должен содержать исходный код, полученные результаты, комментарии, выводы.

**Задание №3: Классификация данных. Неконтролируемая классификация (кластеризация).**

**Цель:** Провести кластеризацию геоданных и классификацию табличных данных.

**Задачи:** для решения задачи кластеризации выбрать набор содержащий геоданные (координаты) (например на сайте Kaggle.com). Провести кластеризацию этих данных (двумя моделями), настроить гиперпараметры, оценить качество кластеризации, визуализировать результат с использованием карты. Для решения задачи классификации выбрать набор табличных данных, обучить три модели, настроить гиперпараметры моделей (использовать grid-search/random-search), оценить качество классификации, построить гос-кривую, выбрать лучшую модель. Сделать выводы по полученным результатам.

**Описание отчета:** Название «№Лабораторной\_Ф\_И\_И\_№группы», отчет в формате html/pdf должен содержать исходный код, полученные результаты, комментарии, выводы.

Критерии оценивания лабораторных работ:

Оценка «зачтено» выставляется, если лабораторные работы выполнены и защищены в полном объеме.

### **3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания**

Видом промежуточной аттестации является зачет с оценкой. Экзаменационный билет состоит из трёх частей, каждая в виде вопроса по одной из тем, освещенной на лекциях (ИОПК-1.1, ИОПК-1.3.1, ИОПК-2.4, ИОПК-3.2, ИОПК-3.3). К зачету с оценкой допускается обучающийся, успешно выполнивший все лабораторные работы, подготовивший и защитивший реферат по теме, согласованной с преподавателем.

Темы лекционных модулей (вопросы для зачета с оценкой):

1. Введение. Основные понятия. Терминология. Области и примеры применения
2. Этапы Data Science
3. Машинное обучение, общая постановка задачи
4. CRISP-DM
5. Регрессия, переобучение
6. Топологии нейросетей и задачи для них
7. Нейросетевая классификация, Deep Learning
8. Сверточные нейронные сети
9. Кластеризация (k-means)
10. Метрики расстояний
11. Критерии точности (Карра, ROC, RMSE), ошибки I/II рода, гипотеза A/B
12. Предварительная обработка данных
13. Оптимизация признакового пространства
14. Классификация (деревья решений)
15. Классификация (статистическая, байесовский подход)
16. SVM (метод опорных векторов)
17. Регуляризация (L1, L2)
18. Ассоциативные алгоритмы (ассоциация, последовательная ассоциация)
19. Высокопроизводительная обработка данных (принципы и модели)
20. Критерии эффективности
21. Многоуровневое машинное обучение. Визуализация
22. Обработка естественного языка
23. Программные среды и сервисы (Hadoop, MapReduce, Spark, Yarn, Cassandra)

Критерии оценивания:

Результаты зачета с оценкой определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» выставляется, если даны правильные ответы на все вопросы без ошибок.

Оценка «хорошо» выставляется, если имеются незначительные неточности в ответах или незначительный дефицит в детализации ответа.

Оценка «удовлетворительно» выставляется, если имеются значительные неточности в ответах или значительный дефицит в детализации ответа.

Оценка «неудовлетворительно» выставляется, если отсутствует понимание предмета.

#### **4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций ИОПК-1.1, ИОПК-1.3.1, ИОПК-2.4, ИОПК-3.2, ИОПК-3.3)**

Проверка остаточных знаний проводится в форме теста.

Примеры вопросов к тесту:

1. Выберите все верные утверждения:
  - а) Последовательность имеет место в случае, если несколько событий связаны друг с другом.
  - б) Отнесение нового объекта к какому-либо из существующих классов выполняется путем классификации.
  - в) В случае, если несколько событий связаны друг с другом во времени, имеет место тип зависимости, именуемый ассоциация.
  - г) Хранимая ретроспективная информация позволяет определить еще одну закономерность, заключающуюся в поиске существующих кластеров.
2. Выявление лояльных или нелояльных держателей кредитных карт относится к задаче
  - а) Контролируемой классификации
  - б) Ассоциации
  - в) Прогнозирования
  - г) Неконтролируемой классификации
3. Для построения алгоритма машинного обучения требуется три типа выборок:
  - а) Обучающая
  - б) Валидационная
  - в) Тестовая
  - г) Стратифицированная
  - д) Квотная
4. Нейросетевые классификаторы относят к:
  - а) Параметрическим подходам
  - б) Непараметрическим подходам
  - в) Прагматическим подходам
  - г) Эклектический подход
5. Индуктивный подход ...
  - а) к исследованию данных позволяет сформулировать гипотезу и найти с ее помощью новые пути аналитических решений
  - б) к исследованию данных предполагает наличие некоторой сформулированной гипотезы, подтверждение или опровержение которой после анализа данных позволяет получить некоторые частные сведения

Ключи: 1 в) 2 г) 3 а)б)в) 4 б) 5 а)

#### **Информация о разработчиках**

Замятин Александр Владимирович, д-р техн. наук, профессор, директор ИПМКН ТГУ.  
Карев Святослав Васильевич, ассистент кафедры теоретических основ информатики ИПМКН ТГУ.