



САМАРСКИЙ УНИВЕРСИТЕТ  
SAMARA UNIVERSITY

# МАГИСТЕРСКАЯ РАБОТА СТУДЕНТА ФИИТ

Самара, 2024



### **Глубокое понимание теоретических основ информационных технологий:**

- владение навыками поиска, критической оценки и правильного использования научных статей, монографий и других литературных источников;
- умение формулировать математическую постановку задачи и гипотезу;
- знание основных подходов проектирования сложных систем;
- изучение и грамотный выбор методов решения задач.

### **Развитие профессиональных компетенций для решения конкретных задач:**

- практические навыки в области: программирования; моделирования; анализа данных; разработки алгоритмов; поиска архитектурных решений и других областей.

### **Проведение самостоятельного исследования, включающее:**

- постановку проблемы;
- формулировку гипотезы;
- сбор и анализ данных;
- разработку и тестирование методов и алгоритмов в различных предметных областях.



## ОБЩАЯ СТРУКТУРА ВКРМ

|                                                                    |                                                                                                                                                                                                                                    |
|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Введение</b>                                                    | должно привлечь внимание к теме исследования и объяснить её актуальность.                                                                                                                                                          |
| <b>Описание предметной области</b>                                 | описание исследуемой проблемы;<br>обоснование актуальности исследования;<br>анализ существующих исследований;<br>обзор систем-аналогов;<br>формулировка цели и задач исследования;<br>формулировка гипотезы исследования.          |
| <b>Методы и модели</b>                                             | описание и обоснование выбора методов и моделей, которые будут использованы для решения конкретной задачи;<br>математическая постановка задачи (при необходимости);<br>описание и выбор метрик оценки качества результатов работы. |
| <b>Разработка системы, как инструмента проведения исследования</b> | проектирование системы: структурная схема, модель базы данных, основные диаграммы UML, экранные формы.                                                                                                                             |
| <b>Исследовательская часть</b>                                     | описание разработанных методов, моделей, алгоритмов, архитектурных решений;<br>демонстрация основных результатов работы, полученных в ходе экспериментов;<br>оценка качества полученных результатов;<br>проверка гипотезы.         |
| <b>Заключение</b>                                                  | обоснование полученных результатов;<br>обсуждение значимости полученных результатов.                                                                                                                                               |





## ВЫБОР И ФОРМУЛИРОВКА ТЕМЫ ИССЛЕДОВАНИЯ

| Характеристика                 | Пример                                                                                                                                                      |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Конкретность</b>            | Вместо широкой темы «Методы машинного обучения» - «Разработка алгоритма классификации космических снимков на основе сверточных нейронных сетей».            |
| <b>Актуальность</b>            | «Оптимизация производительности облачных вычислительных систем» — учитывая рост использования облачных технологий тема является актуальной.                 |
| <b>Новизна</b>                 | «Разработка новой методики защиты данных в сетях Интернета вещей (IoT)» — предполагает исследование и внедрение новых методов защиты данных.                |
| <b>Практическая значимость</b> | «Разработка системы рекомендаций для интернет-магазина на основе поведенческих данных пользователей» — тема с высокой практической значимостью для бизнеса. |

### Как правильно обосновать актуальность темы?

| Характеристика                          | Пример                                                                                                                                                                                                               |
|-----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Ссылки на авторитетные источники</b> | Согласно отчету Gartner (2024), объем данных, генерируемых глобально, увеличивается на 40% ежегодно, что подчеркивает необходимость эффективных методов обработки и анализа данных.                                  |
| <b>Статистические данные</b>            | Согласно исследованию Hootsuite и We Are Social (2023), более 70% пользователей интернета активно используют социальные сети, а платформа ВКонтакте имеет ежемесячную аудиторию более 70 миллионов человек в России. |





| Пример темы                                                                                      | Комментарии                                                                                                                                                                                                                                                                                              |
|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Разработка алгоритма классификации космических снимков на основе сверточных сетей</b>         | <ul style="list-style-type: none"><li>– Конкретность: четко указана область (классификация изображений) и метод (глубокое обучение).</li><li>– Актуальность: важность улучшения алгоритмов классификации для приложений в аэрокосмической отрасли, медицине, безопасности и других областях.</li></ul>   |
| <b>Исследование методов оптимизации маршрутов для беспилотных транспортных средств</b>           | <ul style="list-style-type: none"><li>– Конкретность: определена конкретная задача (оптимизация маршрутов) и область применения (беспилотные транспортные средства).</li><li>– Актуальность: рост интереса к беспилотным транспортным средствам и необходимость их эффективного использования.</li></ul> |
| <b>Разработка системы кибербезопасности для защиты данных в облачных вычислительных системах</b> | <ul style="list-style-type: none"><li>– Конкретность: указана конкретная задача (защита данных) и область применения (облачные вычислительные системы).</li><li>– Актуальность: увеличение использования облачных технологий и необходимость обеспечения их безопасности.</li></ul>                      |



## ТИПЫ ИССЛЕДОВАНИЙ ВКРМ (ФИИТ)

| Тип исследования                                                                                                                                                                                                                | Пример                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Теоретическое исследование</b><br>направлено на разработку новых моделей, алгоритмов или теорий в области информационных технологий. Оно может включать анализ существующих методов и предложений новых решений на их основе | «Разработка нового метода обработки текста для автоматического суммирования новостных статей».<br>Описание: Исследование включает разработку новой архитектуры нейронной сети и её тестирование на открытых наборах данных.                                            |
| <b>Экспериментальное исследование</b><br>предполагает проведение экспериментов для проверки гипотез и оценивания эффективности новых технологий или методов                                                                     | «Сравнительный анализ производительности различных баз данных NoSQL на больших объёмах данных».<br>Описание: Проведение серии экспериментов для оценки скорости чтения и записи данных, масштабируемости и отказоустойчивости различных баз данных NoSQL.              |
| <b>Прикладное исследование</b><br>заключается в решении конкретной практической задачи с использованием ИТ-технологий                                                                                                           | «Разработка системы автоматизированного анализа и классификации медицинских изображений».<br>Описание: Разработка и тестирование системы, способной автоматически классифицировать медицинские изображения (например, МРТ-снимки) с использованием машинного обучения. |
| <b>Комплексное исследование</b><br>может включать элементы теоретического, экспериментального и прикладного подходов для решения многогранной задачи                                                                            | «Создание адаптивной системы онлайн-обучения с элементами искусственного интеллекта».<br>Описание: Исследование включает разработку алгоритмов адаптивного обучения, их интеграцию в онлайн-платформу и экспериментальную проверку эффективности на студентах.         |





## НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ ВКРМ (ФИИТ)

| Направление исследования                                           | Примеры                                                                                                                                                                                                                             |
|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Разработка программного обеспечения для сложных систем и процессов | <ol style="list-style-type: none"><li>1. Разработка мобильного приложения для мониторинга здоровья.</li><li>2. Разработка и оптимизация системы управления контентом (CMS) для малых предприятий.</li></ol>                         |
| Анализ данных                                                      | <ol style="list-style-type: none"><li>1. Анализ и предсказание поведения клиентов на основе исторических данных покупок.</li><li>2. Выявление аномалий в больших данных финансовых транзакций.</li></ol>                            |
| Искусственный интеллект и машинное обучение                        | <ol style="list-style-type: none"><li>1. Анализ поведения пользователей в социальных сетях с помощью методов машинного обучения</li><li>2. Применение глубокого обучения для распознавания образов в медицинских снимках.</li></ol> |
| Моделирование физических процессов                                 | <ol style="list-style-type: none"><li>1. Моделирование распространения эпидемий с использованием агентных моделей.</li><li>2. Компьютерное моделирование гидродинамических процессов в трубопроводах.</li></ol>                     |
| Разработка новых архитектурных решений                             | <ol style="list-style-type: none"><li>1. Оптимизация архитектуры микросервисов для высоконагруженных веб-приложений.</li><li>2. Разработка протокола безопасной передачи данных для устройств IoT.</li></ol>                        |
| Исследование пользовательского опыта (UX/UI)                       | <ol style="list-style-type: none"><li>1. Исследование и улучшение навигации на сайте электронной коммерции.</li><li>2. Разработка интерфейса мобильного приложения для людей с ограниченными возможностями.</li></ol>               |



## ИСТОЧНИКИ ДАННЫХ

| Источники данных                                                                                                                                                                                      |                                                                                                                                                                                                                         |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Публичные базы данных</b><br>предоставляют широкий спектр данных, которые можно использовать для различных исследований. Эти данные часто структурированы и доступны для бесплатного использования | Kaggle – платформа, на которой размещены множество наборов данных для задач машинного обучения и анализа данных. Например, данные для предсказания цен на жилье, анализ медицинских записей и др.                       |
|                                                                                                                                                                                                       | UCI Machine Learning Repository – база данных с набором разнообразных данных для исследований в области машинного обучения, включая данные о винных дегустациях, банковских клиентах и др.                              |
|                                                                                                                                                                                                       | Data.gov – платформа, предоставляющая доступ к данным различных государственных организаций США, включая данные о здоровье, экологии, экономике и др.                                                                   |
| <b>Корпоративные данные,</b><br>предоставленные компаниями, могут быть особенно ценными для исследований, ориентированных на решение бизнес-задач                                                     | Партнёры университета – Самарский университет сотрудничает со множеством компаний, предоставляющими данные для академических исследований. Это могут быть данные о продажах, данные о производственных процессах и т.д. |
|                                                                                                                                                                                                       | Проектные данные – данные, собранные в рамках совместных проектов с корпоративными партнёрами, например, данные о поведении пользователей в онлайн-сервисах.                                                            |
| <b>Результаты экспериментов –</b><br>собираемые в процессе проведения исследований. Позволяют тестировать гипотезы в контролируемых условиях                                                          | Лабораторные эксперименты – данные, собранные в лабораторных условиях, например, результаты тестирования нового алгоритма в контролируемой среде.                                                                       |
|                                                                                                                                                                                                       | Полевые эксперименты – данные, собранные в реальных условиях, например, результаты тестирования нового программного обеспечения в реальной среде эксплуатации.                                                          |
| <b>Симуляционные данные –</b> данные получают в результате моделирования различных процессов и систем. Они полезны, когда реальные данные недоступны или их сбор невозможен                           | Математическое моделирование – данные, полученные в результате математических моделей, например, моделирование расчетов траектории полета космического аппарата или распространения заболеваний.                        |
|                                                                                                                                                                                                       | Компьютерные симуляции – данные, полученные в результате симуляций работы алгоритмов, например, симуляция работы сети при различных нагрузках.                                                                          |



|                               | Применение                                                                                                                                                                                                                                              | Пример                                                                                                                                                                                                 |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Статистический анализ</b>  | Применяется для описательной статистики (среднее, медиана, мода, стандартное отклонение), корреляционного анализа, регрессионного анализа и проверки гипотез.                                                                                           | Анализ результатов опроса пользователей для определения уровня удовлетворенности новым программным обеспечением.<br>Методы: t-тест и ANOVA для проверки значимости различий между группами.            |
| <b>Машинное обучение (ML)</b> | Применяется в задачах классификации, регрессии, кластеризации и рекомендаций. Алгоритмы включают линейную и логистическую регрессию, деревья решений, случайные леса, опорные векторы (SVM) и k-средние.                                                | Разработка модели для предсказания оттока клиентов на основе исторических данных о поведении пользователей.<br>Алгоритмы: случайный лес или градиентный бустинг, для создания предсказательной модели. |
| <b>Глубокое обучение (DL)</b> | Применяется в задачах распознавания образов, обработки естественного языка, генерации текста, машинного перевода и автономных систем. Популярны архитектуры включают свёрточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и трансформеры. | Создание модели для автоматического распознавания изображений медицинских снимков с целью диагностики заболеваний.<br>Свёрточные нейросети для классификации изображений и обнаружения патологий.      |



**Научная гипотеза** в технических науках, особенно в сфере информационных технологий, представляет собой предполагаемое объяснение или решение определённой проблемы или явления, основанное на наблюдениях и предварительных данных.

### **Этапы работы с гипотезой:**

**Формулировка гипотезы** – определение проблемы и предложение гипотетического решения.

**Планирование эксперимента** – определение методов и инструментов, которые будут использованы для проверки гипотезы.

**Проведение эксперимента** – сбор данных и наблюдений в соответствии с планом.

**Анализ результатов** – оценка собранных данных для подтверждения или опровержения гипотезы.

**Выводы** – формулирование выводов на основе анализа результатов и решение о дальнейших действиях.



### Основные характеристики научной гипотезы:

- **Четкость:** ясно указаны алгоритмы и ожидаемый результат.
- **Конкретность:** определены переменные (алгоритмы машинного обучения и традиционные статистические методы) и конкретный показатель (точность прогнозирования).
- **Тестируемость:** можно провести эксперимент, сравнив точность прогнозирования обоих методов на одном и том же наборе данных.
- **Фальсифицируемость:** если данные покажут, что алгоритм машинного обучения X не увеличивает точность на 15%, гипотеза будет опровергнута.

#### ВАЖНО

Правильная формулировка гипотезы помогает сконцентрировать усилия на достижении конкретных целей и обеспечивает основу для структурированного и логически обоснованного исследования.



## НАУЧНАЯ ГИПОТЕЗА

| Область исследования                            | Гипотеза                                                                                                                                                                                            |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Машинное обучение                               | Использование алгоритма градиентного бустинга увеличивает точность предсказания модели на 10% по сравнению с алгоритмом случайного леса при решении задачи классификации спама в электронной почте. |
| Обработка данных                                | Применение метода нормализации данных приведет к снижению средней ошибки предсказания на 10% в модели регрессии.                                                                                    |
| Оптимизация производительности                  | Оптимизация кода с использованием многопоточности уменьшает время выполнения задачи на 30% по сравнению с последовательным выполнением на многопроцессорных системах.                               |
| Управление базами данных                        | Применение индексов в базах данных увеличивает скорость выполнения запросов на 50% по сравнению с базами данных без индексов при выборке большого объёма данных                                     |
| Разработка пользовательского интерфейса (UX/UI) | Использование адаптивного дизайна увеличит удовлетворённость пользователей мобильного приложения на 20% по сравнению с фиксированным дизайном.                                                      |
| Кибербезопасность                               | Внедрение двухфакторной аутентификации уменьшит количество успешных взломов аккаунтов на 50% по сравнению с использованием только пароля.                                                           |



## СПОСОБЫ ПРОВЕРКИ ГИПОТЕЗЫ

| Способ                              | Описание                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>А/В-тестирование</b>             | А/В-тестирование – метод сравнения двух версий продукта или функции (А и В) для определения, какая из них более эффективна. Пользователи случайным образом делятся на две группы: одна группа использует версию А (контрольную), а другая – версию В (тестовую). Сравниваются ключевые метрики (например, конверсия, время на сайте), чтобы определить, какая версия работает лучше.  |
| <b>z-критерий</b>                   | z-критерий используется для проверки гипотез о среднем значении нормального распределения с известной дисперсией. Этот тест сравнивает наблюдаемое значение выборочного среднего с гипотетическим средним, рассчитывая z-статистику, которая показывает, насколько далеко наблюдаемое значение отстоит от среднего значения в терминах стандартных отклонений.                        |
| <b>t-критерий</b>                   | Тест Стьюдента или t-критерий применяется для проверки гипотез о среднем значении, когда дисперсия неизвестна. Он используется для сравнительного анализа двух выборок или для оценки среднего значения одной выборки по отношению к известному значению. t-статистика учитывает размер выборки и стандартное отклонение, что позволяет использовать этот тест для небольших выборок. |
| <b>p-value</b>                      | p-value (уровень значимости) показывает вероятность получения наблюдаемого значения статистики теста или более экстремального значения при условии, что нулевая гипотеза верна. Низкое значение p-value (обычно $< 0.05$ ) указывает на то, что наблюдаемый результат маловероятен.                                                                                                   |
| <b>ANOVA (Analysis of Variance)</b> | ANOVA, или дисперсионный анализ, используется для сравнения средних значений трех и более групп. Цель ANOVA — определить, существуют ли статистически значимые различия между средними значениями этих групп.                                                                                                                                                                         |



Пусть  $X$  – множество объектов,  $Y$  – множество меток классов. Необходимо построить функцию  $f: X \rightarrow Y$ , которая для каждого объекта  $x \in X$  предсказывает метку класса  $y \in Y$ .

**Пример:** Задача классификации электронных писем на «спам» и «не спам».

**Описание:**

- Входные данные (объекты)  $X$ : электронные письма, представленные в виде текстовых данных.
- Метки классов  $Y$ : «спам» и «не спам».
- Функция  $f$ : алгоритм классификации, который определяет, является ли письмо спамом или нет на основе его содержимого.

**Формулировка:**

- $X$  – множество писем
- $Y$  – множество меток классов, где  $Y = \{\text{«спам»}, \text{«не спам»}\}$
- $f(x)=y$ : функция классификации, где  $x$  – это электронное письмо, а  $y$  – метка класса, предсказывающая, является ли письмо спамом или нет.

**Признаки:**

**Частотность определённых слов и фраз:** «бесплатно», «скидка», «победитель» часто встречаются в спам-письмах.

**Наличие ссылок:** спам-письма часто содержат большое количество ссылок.

**Отправитель:** проверка адреса отправителя на соответствие известным спам-адресам.

**Структура письма:** наличие большого количества изображений, мало текста, необычные форматирования.

**Метаданные:** время отправки, заголовки, использование определенных серверов.

$x = (\text{frequency\_free}, \text{number\_of\_links}, \text{known\_spam\_sender}, \text{image\_ratio}, \text{metadata})$

$y$  – метка класса, например, «спам» или «не спам».



## ТИПЫ ЗАДАЧ

|                 | Математическая постановка                                                                                                                                                                               | Пример                                                                                                                                                                                                                                                                        |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Регрессия       | Пусть $X$ – множество входных данных, $Y$ – множество целевых значений.<br>Необходимо построить функцию $f: X \rightarrow Y$ , которая для каждого $x \in X$ предсказывает $y \in Y$ .                  | Задача предсказания уровня дохода.<br>Пусть $x$ – характеристики человека (уровень образования, количество лет опыта работы, возраст и т.д.), а $y$ – уровень дохода. Модель регрессии может быть представлена как: $y = w^T x + b$ , где $w$ – вектор весов, $b$ – смещение. |
| Прогнозирование | Пусть $y_t$ – значение временного ряда в момент времени $t$ . Необходимо построить модель, которая предсказывает будущее значение $y_{t+h}$ на основе прошлых значений $y_t, y_{t-1}, \dots, y_{t-p}$ . | Задача прогнозирования цен акций.<br>Пусть $y_t$ – цена акции в момент времени $t$ . Задача заключается в построении модели, которая предсказывает цену $y_{t+h}$ .                                                                                                           |
| Оптимизация     | Пусть $f(x)$ – целевая функция, необходимо найти $x$ , которое минимизирует или максимизирует $f(x)$ при выполнении ограничений $g_i(x) \leq 0$ .                                                       | Оптимизация маршрутов доставки.<br>Пусть $f(x)$ – общая стоимость доставки, где $x$ – параметры маршрутов. Необходимо найти такие $x$ , которые минимизируют стоимость доставки при выполнении всех временных и логистических ограничений.                                    |
| Кластеризация   | Пусть $X$ – множество объектов, необходимо разбить его на $k$ кластеров $C_1, C_2, \dots, C_k$                                                                                                          | Задача сегментации рынка.<br>Пусть $X$ – множество клиентов с признаками $x_i$ (возраст, доход, покупки и т.д.), задача заключается в нахождении кластеров $C_i$ , которые соответствуют различным сегментам рынка.                                                           |



## ТИПЫ ЗАДАЧ

|                               | Математическая постановка                                                                                                                                                                                                    | Пример                                                                                                                                                                                                                                   |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Распознавание образов         | Пусть $X$ – множество изображений, $Y$ – множество меток классов. Необходимо построить функцию $f: X \rightarrow Y$ , которая для каждого изображения $x \in X$ предсказывает метку класса $y \in Y$ .                       | Задача распознавания рукописных цифр.<br>Пусть $x$ – изображение цифры, задача состоит в построении модели $f(x)$ , которая предсказывает цифру $y$ , изображённую на картинке.                                                          |
| Семантическое сегментирование | Пусть $X$ – изображение, $Y$ – карта меток классов для каждого пикселя изображения. Необходимо построить функцию $f: X \rightarrow Y$ .                                                                                      | Сегментирование медицинских изображений для выделения опухолей.<br>Пусть $x$ – медицинское изображение, $y$ – карта пикселей, где каждый пиксель относится к классу «опухоль» или «норма».                                               |
| Рекомендательные системы      | Пусть $U$ – множество пользователей, $I$ – множество элементов (фильмов, книг и т.д.). Необходимо построить функцию $r: U \times I \rightarrow R$ , которая предсказывает рейтинг $r(u, i)$ элемента $i$ пользователем $u$ . | Задача рекомендаций фильмов.<br>Пусть $u$ – пользователь, $i$ – фильм. Функция $r(u, i)$ предсказывает, насколько пользователю $u$ понравится фильм $i$ на основе его предыдущих оценок.                                                 |
| Снижение размерности          | Пусть $X$ – множество данных высокой размерности, необходимо найти отображение $f: X \rightarrow Z$ , где $Z$ – множество данных низкой размерности, сохраняющее информацию.                                                 | Использование метода главных компонент (PCA) для визуализации высокоразмерных данных.<br>Пусть $X$ – данные с $d$ признаками, необходимо найти $Z$ с меньшим количеством признаков, которое сохраняет как можно больше дисперсии данных. |





## ТИПЫ ЗАДАЧ

|                            | Математическая постановка                                                                                                                                                           | Пример                                                                                                                                                                                                                                      |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Детекция аномалий</b>   | Пусть $X$ – множество наблюдений, необходимо найти подмножество $A \subset X$ , которое содержит аномалии.                                                                          | Детекция мошеннических транзакций в банковской системе. Пусть $x$ – транзакция с её характеристиками (сумма, время, место и т.д.), необходимо определить, является ли транзакция аномальной.                                                |
| <b>Генеративные модели</b> | Пусть $X$ – множество обучающих данных, необходимо обучить модель $p(X)$ , которая способна генерировать новые данные из того же распределения.                                     | Использование генеративных состязательных сетей (GANs) для создания фотореалистичных изображений. Пусть $X$ – набор реальных изображений, задача заключается в обучении модели генерации новых изображений, которые неотличимы от реальных. |
| <b>Предсказание связей</b> | Пусть $G=(V,E)$ – граф, где $V$ – множество вершин (объектов), $E$ – множество рёбер (связей). Необходимо предсказать, появится ли новое ребро $(u,v)$ для пар вершин $u,v \in V$ . | Задача предсказания дружбы в социальной сети. Пусть $V$ – множество пользователей, $E$ – существующие дружеские связи. Задача заключается в предсказании новых дружеских связей $(u,v)$ между пользователями.                               |



## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

|                                                                                                                                                                  | Пример                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Проверка гипотез</b> –<br>сравнение полученных результатов с изначально сформулированными гипотезами с целью подтверждения их                                 | Если гипотеза была, что «алгоритм X улучшит точность классификации изображений на 10% по сравнению с алгоритмом Y», и результаты показывают, что точность действительно возросла на 12%, это подтверждает гипотезу.    |
| <b>Объяснение причин и следствий</b> –<br>проведение исследования, почему разработанный алгоритм или метод дает такие результаты и как это соотносится с теорией | Если алгоритм X показал лучшую производительность, это может быть связано с его способностью лучше обрабатывать определенные типы данных. Необходимо объяснить, какие особенности данных могли повлиять на результаты. |
| <b>Сравнение результатов с предыдущими исследованиями</b> – демонстрация вклада исследования в уже существующие знания                                           | Если предыдущие исследования показывали улучшение точности на 8% при использовании алгоритма Z, а разработанный алгоритм X показал 12%, необходимо сравнить эти результаты и объяснить различия.                       |
| <b>Оценка практической значимости</b> – оценка результатов с практической точки зрения                                                                           | Если алгоритм классификации улучшает точность диагностики заболеваний на 12%, это может существенно повысить эффективность медицинских обследований и уменьшить количество ошибок.                                     |



## ФОРМУЛИРОВАНИЕ ВЫВОДОВ

| Пример                                                                                                                                                                              |                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Обобщение ключевых результатов</b> – подведение итогов основных результатов исследования, краткое изложение, что было обнаружено и как это подтверждает или опровергает гипотезы | Проведенное исследование показало, что алгоритм X улучшает точность классификации изображений на 12% по сравнению с алгоритмом Y, что подтверждает сформулированную гипотезу.                               |
| <b>Выводы и рекомендации</b> – формулировка выводов на основе результатов исследования, предложение рекомендаций для дальнейших исследований или практического применения           | Рекомендуется использовать алгоритм X для задач классификации изображений в медицинских приложениях. В дальнейшем следует исследовать возможности оптимизации этого алгоритма для других типов данных.      |
| <b>Ограничения исследования</b> – формулировка ограничений исследования и обсуждение, как они могли повлиять на результаты. Это демонстрирует объективность и критическое мышление  | Основное ограничение проведенного исследования заключается в ограниченном наборе данных, использованном для тестирования. В будущем необходимо провести тестирование на более разнообразных наборах данных. |
| <b>Перспективы дальнейших исследований</b> – указание направлений для дальнейших исследований, которые могут развить или дополнить результаты                                       | Дальнейшие исследования могут включать разработку гибридных моделей, объединяющих алгоритм X с другими методами, для дальнейшего улучшения точности классификации.                                          |



| Способ визуализации               | Пример                                                                                                         |
|-----------------------------------|----------------------------------------------------------------------------------------------------------------|
| Гистограммы                       | Гистограмма распределения оценок студентов по предмету позволяет увидеть, какие оценки встречаются чаще всего. |
| Диаграммы рассеяния               | Отображение зависимостей между количеством часов подготовки и оценкой на экзамене                              |
| Линейные графики                  | Отображение изменения температуры в течение года                                                               |
| Круговые диаграммы                | Распределение расходов домашнего бюджета по категориям (питание, транспорт, жилье и т.д.)                      |
| Коробчатые диаграммы (box plots)  | Распределение заработных плат в разных отделах компании                                                        |
| Тепловые карты                    | Отображение корреляции между различными характеристиками набора данных                                         |
| Графики плотности (Density Plots) | Плотность распределения времени выполнения задач в системе                                                     |
| Диаграммы потоков (Flow Charts)   | Алгоритм обработки данных в системе                                                                            |



| Способ визуализации                                      | Применение                                                                                                              |
|----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| Обучающие кривые<br>(Learning Curves)                    | Используются для выявления проблем переобучения или недообучения                                                        |
| Confusion Matrix<br>(Матрица ошибок)                     | Используется для оценки качества классификации и выявления конкретных классов, которые модель предсказывает неправильно |
| Receiver Operating Characteristic<br>(ROC)               | Используется для оценки качества классификаторов и выбора оптимального порога                                           |
| Precision-Recall Curve                                   | Используется для оценки производительности моделей в условиях несбалансированных классов                                |
| Тепловые карты активаций<br>(Activation Heatmaps)        | Используется для интерпретации внутренней работы модели и выявления значимых признаков                                  |
| Feature Maps<br>(Карты признаков)                        | Используется для анализа того, как модель извлекает и обрабатывает признаки из данных                                   |
| t-SNE<br>(t-Distributed Stochastic Neighbor Embedding)   | Используется для визуализации распределения данных и кластеров в признаковом пространстве модели                        |
| Gradient-weighted Class Activation Mapping<br>(Grad-CAM) | Используется для интерпретации и объяснения предсказаний модели                                                         |



**Научная статья** – это представление результатов завершённого исследования или его промежуточного этапа.

**Цель:**

- Соблюдение сроков проведения исследования (**срок подачи статьи – не позднее декабря 2 года обучения**)
- Развитие навыков написания научных публикаций у студентов
- Апробация результатов исследований в научном сообществе
- Опыт оформления сопровождающих документов (экспертное заключение)
- Повышение публикационной активности кафедры

| Оценка            | Уровень издания                                |
|-------------------|------------------------------------------------|
| отлично           | журнал из перечня ВАК или индексируемые в RSCI |
| хорошо            | статья в журнале                               |
| удовлетворительно | тезисы доклада                                 |



САМАРСКИЙ УНИВЕРСИТЕТ  
SAMARA UNIVERSITY

# ЖДАНОВА Александра Николаевна

к.т.н., доцент  
доцент кафедры программных  
систем