

ГРНТИ 05.64.54

DOI: <https://doi.org/10.62724/202430308>

**Баигубенова Сая Кубаидоллиевна<sup>1</sup>,**

магистр технических наук, старший преподаватель, НАО «Западно-Казахстанский аграрно-технический университет», Республика Казахстан, 090009, Республика Казахстан, г. Уральск, улица Жангир хана, 51, [saya\\_bsk@mail.ru](mailto:saya_bsk@mail.ru),  
ORCID ID: 0000-0002-2767-9565

**Утешева Гульнара Шаукатовна<sup>2</sup>,**

магистр технических наук, старший преподаватель, ЧВПОУ "Западно-Казахстанский инновационно-технологический университет", Республика Казахстан, 090006, г.Уральск, улица М.Ихсанова 44/1, [utesheva.gulnara@mail.ru](mailto:utesheva.gulnara@mail.ru),  
ORCID ID: 0000-0003-2253-7493

## ТЕОРИЯ НЕЧЕТКИХ МНОЖЕСТВ В ИНТЕЛЛЕКТУАЛЬНОМ АНАЛИЗЕ ДАННЫХ

**Аннотация:** Статья посвящена применению интеллектуального анализа данных (KDD) и нечетких множеств для работы с большими объемами данных. Цифровая революция привела к значительному увеличению объема и скорости данных, что делает традиционные методы анализа неэффективными. Интеллектуальный анализ данных охватывает этапы очистки, интеграции, преобразования и извлечения знаний из данных. Нечеткая логика, расширяющая булеву логику, позволяет учитывать неопределенность и приближенные рассуждения, что делает её особенно полезной для обработки сложных и зашумленных данных. В статье подробно рассматриваются применения нечетких множеств в таких задачах, как кластеризация, обнаружение ассоциативных правил, обобщение данных, анализ временных рядов и веб-приложения. Нечеткая логика помогает решать проблемы, связанные с неполными или неточными данными, предоставляя более гибкие и понятные решения. Автор также отмечает, что использование нечетких множеств улучшает адаптацию систем к специфическим требованиям и особенностям данных, что способствует более точному анализу и принятию решений в таких областях, как здравоохранение, финансы и телекоммуникации. Эти преимущества делают нечеткие множества важным инструментом для повышения эффективности интеллектуального анализа данных и улучшения взаимодействия с пользователем.

**Ключевые слова:** Интеллектуальный анализ данных, нечеткое множество, обнаружение знаний на основе данных (KDD).

### I. ВВЕДЕНИЕ

Цифровая революция сделала оцифрованную информацию простой в получении и относительно недорогой в хранении [1], [2]. С развитием компьютерного оборудования и программного обеспечения и стремительной компьютеризацией бизнеса в базах данных было собрано и сохранено огромное количество данных. Скорость, с которой хранятся такие данные, растет с феноменальной скоростью. В результате традиционные специальные комбинации статистических методов и инструментов управления данными больше не подходят для анализа этого обширного массива данных. К числу областей, в которых в централизованных или распределенных базах данных хранятся большие объемы данных, относятся следующие.

- Финансовые инвестиции: фондовые индексы и цены, процентные ставки, данные кредитных карт, обнаружение мошенничества [3].
- Здравоохранение: Некоторые диагностические данные, хранящиеся в системах управления больницами [4].
- Производство: оптимизация процессов и устранение неполадок [5].
- Телекоммуникационные сети: схемы вызовов и системы управления неисправностями.
- Научная область: Астрономические наблюдения [6], геномные данные, биологические данные.
- Всемирная паутина [7].

Интеллектуальный анализ данных (data mining) - это извлечение знаний из больших массивов данных. Интеллектуальный анализ - это яркий термин, характеризующий процесс, при котором из большого количества сырья извлекается небольшой набор ценных самородков. Таким образом, такое неправильное название, которое включает в себя как “данные”, так и “интеллектуальный анализ”, стало популярным. Многие другие термины имеют схожее или слегка отличающееся значение с интеллектуальным анализом данных, такие как извлечение знаний из данных, извлечение знаний, анализ данных/шаблонов, археология данных и углубление в данные.

Нечеткая логика быстро стала одной из самых успешных на сегодняшний день технологий для разработки сложных систем управления. Причина этого очень проста. Нечеткая логика идеально подходит для таких приложений, поскольку она напоминает процесс принятия решений человеком и позволяет генерировать точные решения на основе определенной или приблизительной информации. В то время как другие подходы требуют точных уравнений для моделирования поведения в реальном мире, нечеткий дизайн может учесть неоднозначности реального мира в человеческом языке и логике. Хотя генетические алгоритмы и нейронные сети во многих случаях могут работать так же хорошо, как и нечеткая логика, нечеткая логика имеет то преимущество, что решение проблемы может быть сформулировано в понятных операторам терминах, так что их опыт может быть использован при проектировании системы.

контроллер. Это упрощает механизацию задач, которые уже успешно выполняются людьми. В настоящей статье представлен обзор доступной и весьма скудной литературы по интеллектуальному анализу данных в нечетком множестве. В разделе II описаны основные понятия обнаружения знаний в базах данных и интеллектуального анализа данных. Освещены некоторые проблемы. В разделе III приводится обзор, объясняющий роль нечеткого множества. Раздел IV завершает статью.

**II. KDD и ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ** Многие люди рассматривают интеллектуальный анализ данных как синоним другого широко используемого термина - обнаружение знаний на основе данных, или KDD. Тематика KDD развивалась и продолжает развиваться на стыке исследований в таких областях, как базы данных, машинное обучение, распознавание образов, статистика, искусственный интеллект, рассуждения в условиях неопределенности, получение знаний для экспертных систем, визуализация данных, машинное обнаружение и высокопроизводительные вычисления. Системы KDD включают в себя теории, алгоритмы и методы из всех этих областей. Сообщалось о множестве успешных применений в различных секторах, таких как маркетинг, финансы, банковское дело, производство и телекоммуникации.

Теории и инструменты баз данных обеспечивают необходимую инфраструктуру для хранения, доступа к данным и манипулирования ими. Недавно ставший популярным

термин "хранилище данных" [2] относится к современным бизнес-тенденциям в области сбора и очистки транзакционных данных и предоставления их в распоряжение для анализа и поддержки принятия решений.

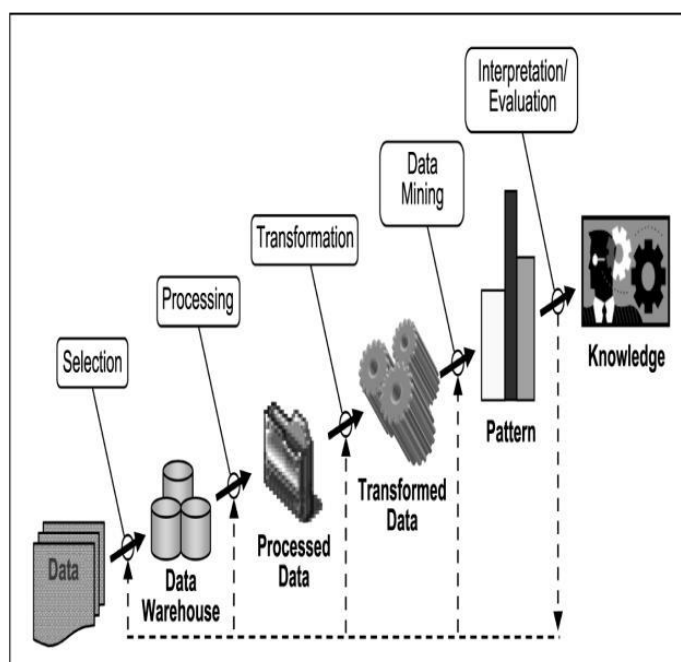


Рисунок 1 Процесс KDD

Процесс обнаружения знаний показан на рисунке 1 и состоит из итеративной последовательности следующих шагов:

1. Очистка данных (для удаления шума и несогласованных данных)
2. Интеграция данных (при которой можно объединить несколько источников данных)
3. Отбор данных (когда данные, относящиеся к задаче анализа, извлекаются из базы данных)
4. Преобразование данных (когда данные преобразуются или консолидируются в формы, подходящие для интеллектуального анализа, например, путем выполнения операций обобщения или агрегирования)
5. Интеллектуальный анализ данных (важный процесс, при котором применяются интеллектуальные методы для извлечения шаблонов данных)
6. Оценка шаблонов (для выявления действительно интересных шаблонов, представляющих знания, на основе некоторых показателей интересности)
7. Представление знаний (при котором используются методы визуализации и представления знаний для представления полученных знаний пользователю).

Интеллектуальный анализ данных включает в себя подгонку моделей к наблюдаемым данным или определение закономерностей на их основе. Подогнанные модели играют роль предполагаемых знаний. Решение о том, отражает ли модель полезные знания или нет, является частью общего процесса KDD, для которого обычно требуется субъективное суждение человека. Как правило, алгоритм интеллектуального анализа данных представляет собой некоторую комбинацию из следующих трех компонентов. • Модель: функция модели (например, классификация, кластеризация) и ее форма представления (например, линейные дискриминанты, нейронные сети). Модель содержит параметры, которые должны быть определены на основе данных. • Критерий

предпочтения: основа для предпочтения одной модели или набора параметров другой в зависимости от заданных данных. Критерием обычно является некоторая форма соответствия модели данным, возможно, смягченная сглаживающим фактором, чтобы избежать чрезмерной подгонки, или создание модели со слишком большим количеством степеней свободы, чтобы ее можно было ограничить заданными данными.

• Алгоритм поиска: спецификация алгоритма для поиска конкретных моделей и параметров с учетом данных, моделей(ов) и критерия предпочтения. Конкретный алгоритм интеллектуального анализа данных обычно представляет собой реализацию компонентов модели/предпочтений/поиска. Наиболее распространенные функции модели в современной практике интеллектуального анализа данных включают следующее.

1) Классификация: относит элемент данных к одному из нескольких predeterminedных категориальных классов.

2) Регрессия: сопоставляет элемент данных с прогнозируемой переменной с реальным значением.

3) Кластеризация: сопоставляет элемент данных с одним из нескольких кластеров, где кластеры представляют собой естественные группировки элементов данных на основе показателей сходства или моделей плотности вероятности.

4) Генерация правил: извлекает правила классификации из данных.

5) Обнаружение ассоциативных правил: описывает ассоциативные отношения между различными атрибутами.

6) Обобщение: предоставляет краткое описание для подмножества данных.

7) Моделирование зависимостей: описывает существенные зависимости между переменными.

Топ-10 сложных проблем в исследованиях интеллектуального анализа данных:

1) Разработка объединяющей теории интеллектуального анализа данных: Современное состояние исследований в области интеллектуального анализа данных является слишком “разрозненным”, поэтому необходимы объединяющие исследования.

2) Расширение масштабирования для данных с высокой размерностью и высокоскоростных потоков: задачи классификации с высокой размерностью (миллионы или миллиарды объектов, например, биоданные) и сверхскоростные потоки данных.

3) Последовательные данные и временные ряды: как эффективно и точно группировать, классифицировать и прогнозировать тенденции?

4) Извлечение сложных знаний из сложных данных: многие объекты не являются независимыми друг от друга и не относятся к одному типу. Изучите богатую структуру связей между объектами. Например, взаимосвязанные веб-страницы, социальные сети, метаболические сети в клетке.

5) Интеллектуальный анализ данных в сетевых условиях: Взаимосвязанные данные между электронными письмами, веб-страницами, блогами, цитатами, последовательностями и людьми.

6) Распределенный интеллектуальный анализ данных и сбор многоагентных данных: необходимо сопоставлять данные, полученные при различных исследованиях.

7) Интеллектуальный анализ данных для решения биологических и экологических проблем: Интеллектуальный анализ биологических данных, таких как разработка вакцины против ВИЧ, ДНК, химические свойства, трехмерные структуры и функциональные свойства -> необходимо объединить, интеллектуальный анализ экологических данных, интеллектуальный анализ для решения энергетического кризиса.

8) Интеллектуальный анализ данных - Проблемы, связанные с процессом: состав операций интеллектуального анализа данных, очистка данных с возможностью ведения журнала, визуализация и автоматизация интеллектуального анализа.

9) Безопасность, конфиденциальность и целостность данных: Как обеспечить конфиденциальность пользователей во время сбора их данных?

10) Работа с нестатическими, несбалансированными и чувствительными к затратам данными: Реальные данные очень велики ( $10^5$  функций), но содержат только  $< 1\%$  полезных классов. [8]

### III. НЕЧЕТКИЕ МНОЖЕСТВА

Моделирование неточных и качественных знаний, а также передача и обработка неопределенности на различных этапах возможны благодаря использованию нечетких множеств. Нечеткая логика способна в разумных пределах поддерживать рассуждения человеческого типа в естественной форме. Это самая ранняя и наиболее широко используемая составляющая мягких вычислений. Нечеткая логика - это надмножество обычной (булевой) логики, которая была расширена для обработки концепции частичной истинности - значений истинности между "полностью истинным" и "полностью ложным". Как следует из названия, именно логика, лежащая в основе способов рассуждения, является скорее приблизительной, чем точной. Важность нечеткой логики обусловлена тем фактом, что большинство способов человеческого мышления и особенно рассуждения, основанные на здравом смысле, являются приблизительными по своей природе. Основные характеристики нечеткой логики, основанной Зайдером Лотфи, заключаются в следующем.

- В нечеткой логике точное рассуждение рассматривается как предельный случай приближенного рассуждения.
- В нечеткой логике все зависит от степени.
- Любая логическая система может быть фаззифицирована.
- В нечеткой логике знание интерпретируется как совокупность эластичных или, что эквивалентно, нечетких ограничений на набор переменных.
- Логический вывод рассматривается как процесс распространения эластичных ограничений.

Нечеткие множества - это обобщенные множества, которые допускают градуированную принадлежность своих элементов. Обычно в качестве структуры степени принадлежности выбирается реальный единичный интервал  $[0; 1]$ . Пусть  $X$  - пространство точек, общий элемент которого обозначается через  $x$ . Таким образом,  $X = \{x\}$ . Нечеткое множество  $A$  в  $X$  характеризуется функцией принадлежности  $f_A(x)$ , которая связывает с каждой точкой в  $X$  действительное число в интервале  $[0,1]$ , причем значения  $f_A(x)$  в точке  $x$  представляют "степень принадлежности"  $x$  в  $A$ . Таким образом, чем ближе значение от  $f_A(x)$  до единицы, тем выше степень принадлежности  $x$  к  $A$ .

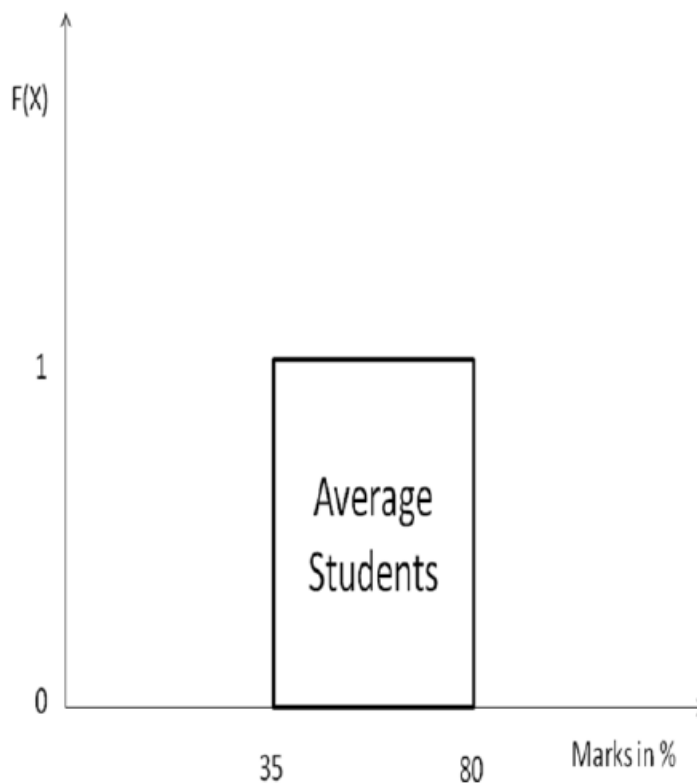
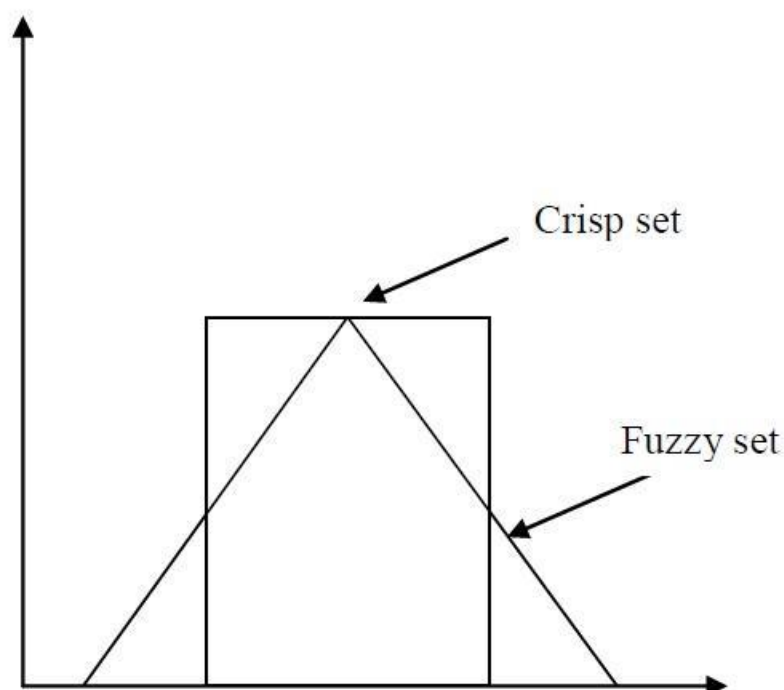


Рисунок 2 Пример задачи с четкой границей



Предположим, у нас есть три диапазона оценок по любому экзамену.  $F(x)$  - это такая функция, что  $0 < X \leq 35$ , тогда  $f(X)$  = ученики-неудачники  $35 < X \leq 80$ , тогда  $f(X)$  = средние ученики  $X > 80$ , тогда  $f(X)$  = умные ученики Из приведенного выше сценария

теперь предположим, что студент набрал 79,8% баллов. значит, он средний ученик. Но если он набрал 80,1 балла, то будет присвоен тег "умный". Но ученик, набравший 79,8 % баллов, также является умным учеником, что не является существенной разницей между обоими условиями. На рисунке 2 показан сценарий для задачи о форме границы, который был описан выше.

Существует несколько основных подходов к решению задачи с четкими границами.

1. Количественный подход
2. Нечеткие таксономические структуры
3. Приближенный подход к набору элементов

Для решения задачи с четкими границами с использованием количественного подхода разделите переменные метки на три нечетких множества. Нечеткие множества и их функции принадлежности должны быть определены экспертом в предметной области. Для простоты демонстрации мы просто определим границы множеств и разделим перекрывающуюся часть поровну между сгенерированными таким образом нечеткими множествами. В качестве примера мы будем использовать следующие границы для нечетких множеств переменных оценок: Неуспеваемость= $\{0-35\}$ , Среднестатистические учащиеся= $\{33-70\}$ , интеллектуальные учащиеся= $\{70-\infty\}$ . Сгенерированные нечеткие множества показаны на рисунке 2. Для всех областей, в которых наборы не перекрываются, поддержка будет просто равна 1 для фактического набора элементов. Если есть перекрытие, принадлежность может быть вычислена с использованием границ перекрывающихся нечетких множеств. Добавленная поддержка в данном случае всегда будет равна 1. Теория нечетких множеств все чаще используется в интеллектуальных системах из-за ее простоты и сходства с человеческим мышлением. Неоспоримо возрастает роль технологии нечетких множеств в области интеллектуального анализа данных [9]. Различные браузеры данных были реализованы с использованием теории нечетких множеств [10]. Анализ реальных данных при интеллектуальном анализе данных часто требует одновременной работы с различными типами переменных, а именно с категориальными/символьными данными и числовыми данными. Наукк [11] разработал алгоритм обучения, который создает смешанные нечеткие правила, включающие как категориальные, так и числовые атрибуты. Педрич [12] обсуждает некоторые конструктивные и управляемые нечеткими множествами вычислительные механизмы обнаружения знаний и устанавливает взаимосвязь между интеллектуальным анализом данных и нечетким моделированием. Ниже приведена классификация нечетких множеств на основе различных функций интеллектуального анализа данных, которые моделируются.

1) Кластеризация: Интеллектуальный анализ данных направлен на просеивание больших объемов данных с целью выявления полезной информации в виде новых взаимосвязей, шаблонов или кластеров для принятия решений пользователем [13].

2) Ассоциативные правила: Важная область исследований интеллектуального анализа данных связана с выявлением ассоциативных правил [14]. Ассоциативное правило описывает интересную ассоциативную связь между различными атрибутами.

3) Функциональные зависимости: Нечеткая логика используется для анализа логических выводов, основанных на функциональных зависимостях (FD) между переменными в связях с базой данных. Нечеткий вывод обобщает как неточный (многозначный), так и точный вывод. Аналогичным образом, нечеткие реляционные базы данных обобщают свои классические и неточные аналоги, поддерживая хранение и поиск нечеткой информации [15].

4) Обобщение данных: Поиск кратких сведений является одним из основных компонентов поиска знаний в базах данных. Это предоставляет пользователю исчерпывающую информацию, позволяющую извлечь суть из большого объема информации в базе данных. Теория нечетких множеств также используется для обобщения данных [16].

5) Веб-приложение: Извлечение типичных профилей пользователей и ассоциаций URL-адресов из огромного количества журналов доступа является важным компонентом веб-персонализации, который связан с адаптацией взаимодействия пользователя с информационным веб-пространством на основе информации о нем / ней. Насрауи и др. [17] определили пользовательскую сессию как компактную во времени последовательность веб-обращений пользователя и использовали показатель различия между двумя веб-сессиями для определения организации веб-сайта. Их цель - классифицировать эти сеансы с помощью веб-майнинга.

6) Поиск изображений: Недавнее увеличение объема хранилищ мультимедийной информации, состоящих из смешанных мультимедийных данных, сделало поиск изображений на основе контента (CBIR) активной исследовательской областью [18]. Нечеткое множество, используемое в химической, медицинской, образовательной и многих других областях.

#### IV. ЗАКЛЮЧЕНИЕ

Современные исследования в области интеллектуального анализа данных в основном сосредоточены на алгоритмах обнаружения и методах визуализации. Растет понимание того, что на практике в базе данных легко обнаружить огромное количество шаблонов, большинство из которых на самом деле очевидны, избыточны и бесполезны или неинтересны пользователю. Чтобы пользователь не был перегружен большим количеством неинтересных шаблонов, необходимы методы, позволяющие выявлять только полезные/ интересные шаблоны и представлять их пользователю. Нечеткие множества, которые являются старейшим компонентом мягких вычислений, подходят для решения проблем, связанных с понятностью шаблонов, неполными/зашумленными данными, смешанной информацией и взаимодействием с человеком, и могут быстрее предоставлять приблизительные решения. В основном они использовались для кластеризации, обнаружения ассоциативных правил и функциональных зависимостей, обобщения, анализа временных рядов, веб-приложений и поиска изображений.

#### СПИСОК ЛИТЕРАТУРЫ

1. Иванов, И. А. Интеллектуальный анализ данных: Теория и практика [Текст] / И. А. Иванов. — Москва: Научный мир, 2018.
2. Петров, В. В., & Сидоров, А. Б. Методы интеллектуального анализа данных: Основы и приложения [Текст] / В. В. Петров, А. Б. Сидоров. — Санкт-Петербург: Лань, 2020.
3. Лебедев, Н. Н. Нечеткие множества и их применение в интеллектуальном анализе данных [Текст] / Н. Н. Лебедев. — Москва: ИТД, 2017.
4. Кузнецов, А. Ю. Введение в нечеткую логику и ее применение в аналитике данных [Текст] / А. Ю. Кузнецов. — Новосибирск: Сибирское издательство, 2019.
5. Романов, В. Г. Нечеткие множества в вычислительных системах и их использование в анализе данных [Текст] / В. Г. Романов. — Екатеринбург: УрФУ, 2021.
6. Смирнова, И. В. Моделирование и анализ данных с помощью нечетких множеств [Текст] / И. В. Смирнова. — Казань: Казанский университет, 2018.

7. Алексеева, М. Д. Теория и методы нечетких множеств: Применение в области анализа данных [Текст] / М. Д. Алексеева. — Москва: Физматлит, 2022.
8. Гусев, И. В. Анализ данных с использованием методов нечеткой логики [Текст] / И. В. Гусев. — Ростов-на-Дону: ЮФУ, 2020.
9. Дмитриев, С. В. Интеллектуальный анализ данных и нечеткие множества: Современные тенденции [Текст] / С. В. Дмитриев. — Владивосток: ДВФУ, 2019.
10. Казакова, О. А. Нечеткие множества в области интеллектуального анализа данных и их алгоритмическое обеспечение [Текст] / О. А. Казакова. — Томск: Томский университет, 2021.
11. Наукк, Д. Использование символьных данных в нейро-нечеткой классификации [Текст] / Д. Наукк. — В сборнике: NAFIPS 99, Нью-Йорк, июнь 1999 г., стр. 536-540.
12. Педрич, У. Технология нечетких множеств в поиске знаний [Текст] / У. Педрич. — Система нечетких множеств, том 98, стр. 279-290, 1998.
13. Пятецкий-Шапиро, П., & Фроули, У. Дж. (Ред.). Открытие знаний в базах данных [Текст] / П. Пятецкий-Шапиро, У. Дж. Фроули (Ред.). — Менло-Парк, Калифорния: AAAI/MIT Press, 1991.
14. Агравал, Р., Имиелински, Т., & Свами, А. Правила поиска ассоциаций между наборами элементов в больших базах данных [Текст] / Р. Агравал, Т. Имиелински, А. Свами. — В сборнике: 1993 ACM SIGMOD Int. Conf. Управленческие данные, Вашингтон, округ Колумбия, май 1993 г., стр. 207-216.
15. Хейл, Дж., & Шенои, С. Анализ логического вывода в реляционных базах данных [Текст] / Дж. Хейл, С. Шенои. — Data Knowledge Eng., том 18, стр. 167-183, 1996.
16. Ли, Д. Х., & Ким, М. Х. Обобщение базы данных с использованием нечетких иерархий ISA [Текст] / Д. Х. Ли, М. Х. Ким. — IEEE Trans. Система, человек, кибернетика. В, том 27, стр. 68-78, 1997.
17. Насрауи, О., Кришнапурам, Р., & Джоши, А. Реляционная кластеризация на основе нового надежного метода оценки с применением к веб-майнингу [Текст] / О. Насрауи, Р. Кришнапурам, А. Джоши. — В Proc. NAFIPS 99, Нью-Йорк, июнь 1999 г., стр. 705-709.
18. Пал, С. К., Гош, А., & Кунду, М. К. (Ред.). Программные вычисления для обработки изображений [Текст] / С. К. Пал, А. Гош, М. К. Кунду (Ред.). — Гейдельберг, Германия: Physica-Verlag, 2000.
19. Митра, С., & Митра, П. Интеллектуальный анализ данных в среде мягких вычислений: Обзор [Текст] / С. Митра, П. Митра. — IEEE TRANSACTIONS ON NEURAL NETWORKS, том 13, № 1, стр. 3-14, январь 2002.
20. Хан, Цзявэй, & Камбер, Мишлин. Концепции и методы интеллектуального анализа данных [Текст] / Цзявэй Хан, Мишлин Камбер. — 2-е издание.

## REFERENCES

1. Ivanov, I. A. "Intellectual'nyj analiz dannyh: Teoriya i praktika [Data Mining: Theory and Practice]." Moskva, Nauchnyj mir, (2018): - (In Rus)
2. Petrov, V. V., & Sidorov, A. B. Metody intellektual'nogo analiza dannyh: Osnovy i prilozheniya [Data Mining Methods: Fundamentals and Applications]." Sankt-Peterburg, Lan', (2020): - (In Rus)
3. Lebedev, N. N. Nечеткие mnozhestva i ih primeneniye v intellektual'nom analize dannyh [Fuzzy sets and their application in data mining]." Moskva, ITD, (2017): - (In Rus)

4. Kuznecov, A. YU. Vvedenie v nechetkuyu logiku i ee primenenie v analitike dannyh [Introduction to Fuzzy Logic and its application in data analytics].” Novosibirsk, Sibirskoe izdatel'stvo, (2019): - (In Rus)
5. Romanov, V. G. Nechetkie mnozhestva v vychislitel'nyh sistemah i ih ispol'zovanie v analize dannyh [Fuzzy sets in computing systems and their use in data analysis].” Ekaterinburg, UrFU, (2021): - (In Rus)
6. Smirnova, I. V. Modelirovanie i analiz dannyh s pomoshch'yu nechetkih mnozhestv [Modeling and data analysis using fuzzy sets].”Kazan. Kazanskij universitet, (2018): - (In Rus)
7. Alekseeva, M. D. Teoriya i metody nechetkih mnozhestv: Primenenie v oblasti analiza dannyh [. Theory and methods of fuzzy sets: Application in the field of data analysis].” Moskva, Fizmatlit, (2022): - (In Rus)
8. Gusev, I. V. Analiz dannyh s ispol'zovaniem metodov nechetkoj logiki [Data analysis using fuzzy logic methods].” Rostov-na-Donu, YUFU, (2020): - (In Rus)
9. Dmitriev, S. V. Intel'ektual'nyj analiz dannyh i nechetkie mnozhestva: Sovremennye tendencii [Data mining and fuzzy sets: Current trends].” Vladivostok: DVFU, (2019): - (In Rus)
10. Kazakova, O. A. Nechetkie mnozhestva v oblasti intellektual'nogo analiza dannyh i ih algoritmicheskoe obespechenie [Fuzzy sets in the field of data mining and their algorithmic support].” Tomsk, Tomskij universitet, (2021): - (In Rus)
11. Naukk, D. Ispol'zovanie simvol'nyh dannyh v nejro-nechetkoj klassifikacii [The use of symbolic data in neuro-fuzzy classification].” V sbornike: NAFIPS 99, N'yu-Jork, iyun' str. 536-540, (1999): - (In Kaz)
12. Pedrich, U. Tekhnologiya nechetkih mnozhestv v poiske znaniy [The technology of fuzzy sets in the search for knowledge].” Sistema nechetkih mnozhestv, tom 98, str. 279-290, (1998): - (In Rus)
13. Pyateckij-SHapiro, P., & Frouli, U. Dzh. (Red.). Otkrytie znaniy v bazah dannyh [Discovery of knowledge in databases].” Menlo-Park, Kaliforniya, AAAI/MIT Press, (1991): - (In Rus)
14. Agraval, R., Imielinski, T., & Svami, A. Pravila poiska asociacij mezhdu naborami elementov v bol'shih bazah dannyh [Rules for finding associations between sets of elements in large databases].”V sbornike: 1993 ACM SIGMOD Int. Conf. Upravlencheskie dannye, Vashington, okrug Kolumbiya, maj g., str. 207-216, (1993): - (In Rus)
15. Hejl, Dzh., & SHenoi, S. Analiz logicheskogo vyvoda v relyacionnyh bazah dannyh [Logical inference analysis in relational databases].” Data Knowledge Eng., tom 18, str. 167-183, (1996): - (In Rus)
16. Li, D. H., & Kim, M. H. Obobshchenie bazy dannyh s ispol'zovaniem nechetkih ierarhij ISA [Generalizing the database using fuzzy ISA hierarchies].” IEEE Trans. Sistema, chelovek, kibernetika. V, tom 27, str. 68-78, (1997): - (In Rus)
17. Nasraui, O., Krishnapuram, R., & Dzhoshi, A. Relyacionnaya klasterizaciya na osnove novogo nadezhnogo metoda ocenki s primeneniem k veb-majningu [Relational clustering based on a new reliable estimation method applied to web mining].”V Proc. NAFIPS 99, N'yu-Jork, iyun' g., str. 705-709, (1999): - (In Rus)
18. Pal, S. K., Gosh, A., & Kundu, M. K. (Red.). Programmnye vychisleniya dlya obrabotki izobrazhenij [Software calculations for image processing].” Gejdel'berg, Germaniya: Physica-Verlag, (2000): - (In Rus)
19. Mitra, S., & Mitra, P. Intel'ektual'nyj analiz dannyh v srede myagkih vychislenij: Obzor [Data Mining in a Soft Computing environment: An Overview].” IEEE

TRANSACTIONS ON NEURAL NETWORKS, tom 13, № 1, str. 3-14, yanvar' (2002):- (In Rus)

20. Han, Czyavej, & Kamber, Mishlin. *Koncepcii i metody intellektual'nogo analiza dannyh [Concepts and methods of data mining].* 2-e izdanie: - (In Rus)

## ДЕРЕКТЕРДІ ӨНДІРУДЕГІ АНЫҚ ЕМЕС ЖИЫНТЫҚТАР ТЕОРИЯСЫ

**Аннотация:** Мақала үлкен көлемдегі деректермен жұмыс істеу үшін деректерді өндіруді (KDD) және анық емес жиындарды қолдануға арналған. Сандық революция деректердің көлемі мен жылдамдығының айтарлықтай өсуіне әкелді, бұл дәстүрлі талдау әдістерін тиімсіз етеді. Деректерді өндіру деректерді тазарту, біріктіру, түрлендіру және деректерді алу қадамдарын қамтиды. Логикалық логиканы кеңейтетін анық емес логика белгісіздік пен болжамды пайымдауды ескеруге мүмкіндік береді, бұл оны күрделі және шулы деректерді өңдеу үшін әсіресе пайдалы етеді. Мақалада кластерлеу, ассоциативті ережелерді анықтау, деректерді жалпылау, уақыт қатарын талдау және веб-қосымшалар сияқты тапсырмаларда анық емес жиындардың қолданылуы егжей-тегжейлі қарастырылады. Анық емес логика икемді және түсінікті шешімдер ұсына отырып, толық емес немесе дәл емес деректерге қатысты мәселелерді шешуге көмектеседі. Сондай-ақ, Автор анық емес жиындарды пайдалану жүйелердің нақты талаптар мен деректер ерекшеліктеріне бейімделуін жақсартады, бұл денсаулық сақтау, қаржы және телекоммуникация сияқты салаларда дәлірек талдау мен шешім қабылдауға ықпал етеді. Бұл артықшылықтар анық емес жиындарды деректерді өндіру тиімділігін арттырудың және пайдаланушы тәжірибесін жақсартудың маңызды құралына айналдырады.

**Кілт сөздер:** деректерді өндіру, анық емес жиын, деректерге негізделген білімді анықтау (KDD).

## FUZZY SET THEORY IN DATA MINING

**Abstract:** The article is devoted to the application of data mining (KDD) and fuzzy sets to work with large amounts of data. The digital revolution has led to a significant increase in the volume and speed of data, which makes traditional analysis methods ineffective. Data mining covers the stages of purification, integration, transformation and extraction of knowledge from data. Fuzzy logic, which extends Boolean logic, allows for uncertainty and approximate reasoning, which makes it especially useful for processing complex and noisy data. The article discusses in detail the applications of fuzzy sets in tasks such as clustering, associative rule detection, data generalization, time series analysis, and web applications. Fuzzy logic helps solve problems related to incomplete or inaccurate data by providing more flexible and understandable solutions. The author also notes that the use of fuzzy sets improves the adaptation of systems to specific data requirements and features, which contributes to more accurate analysis and decision-making in areas such as healthcare, finance and telecommunications. These advantages make fuzzy sets an important tool for improving the efficiency of data mining and improving user interaction.

**Keywords:** Data mining, fuzzy set, data-based knowledge discovery (KDD).