

МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра компьютерной физики и метаматериалов  
на базе Саратовского филиала Института радиотехники  
и электроники им. В. А. Котельникова РАН

## **МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

### **АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

Максимова Кирилла Владимировича,  
студента 4 курса, группа 4022,  
направления подготовки 03.03.02 Физика,  
Институт физики

Научный руководитель  
д.ф.-м.н. профессор

О.А. Черкасова

Заведующий кафедрой  
компьютерной физики и метаматериалов  
на базе Саратовского филиала Института радиотехники  
и электроники им. В. А. Котельникова РАН  
д.ф.-м.н. профессор

В.М. Аникин

Саратов 2021 г.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** В связи с интенсивным развитием информационных технологий, увеличивается и создаваемая ими информация, в результате чего большая её часть просто физически не может быть обработана человеком, следовательно, значительные объёмы практичной и полезной информации просто теряются. На сегодняшний момент технологии Data Mining широко используются во многих областях науки – биологии, медицине, астрономии и бизнеса – в торговле, телекоммуникации, промышленном производстве и других.

**Целью данной работы** является изучение существующих методов интеллектуального анализа данных и исследование экспериментальных данных при помощи приложения «Анализ данных» MS Excel.

**Задачами работы** являются:

- 1) изучить теоретические основы интеллектуального анализа данных;
- 2) рассмотреть существующие методы интеллектуального анализа;
- 3) применить методы интеллектуального анализа данных в MS Excel.

**Объект исследования:** Data Mining — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Data Mining представляют большую ценность для руководителей и аналитиков в их повседневной деятельности, т.к. подходят для аналитики больших объемов данных.

**Структура ВКР.** Выпускная квалификационная работа (ВКР) представлена на 68 страницах. Структура работы включает: введение, 3 главы, заключение, список использованных источников (122 наименования, в том числе 65 статей в периодических изданиях за последние 20 лет и 30 – иностранных авторов). В тексте имеется 22 рисунка и 1 таблица.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Актуальность темы.** В связи с интенсивным развитием информационных технологий, увеличивается и создаваемая ими информация, в результате чего большая её часть просто физически не может быть обработана человеком, следовательно, значительные объёмы практичной и полезной информации просто теряются. Кроме того, важные решения могут приниматься не на основе анализа данных, накопленных в информационных базах данных, а на абстрактном механизме – интуиции человека. Это приводит к пониманию важности проблемы анализа данных, полученных искусственными интеллектуальными системами. Следовательно, для извлечения нужной информации в огромном океане данных требуется использование методов интеллектуального анализа данных.

На сегодняшний момент технологии Data Mining широко используются во многих областях науки – биологии, медицине, астрономии и бизнеса – в торговле, телекоммуникации, промышленном производстве и других. На рынке информаци-

онных технологий представлено большое количество платформ, использующих технологии интеллектуального анализа данных, коммерческого и свободного распространения.

На основании вышеизложенного можно заключить об актуальности исследования методов интеллектуального анализа данных.

В главе 1 рассматриваются сферы применения технологий Data Mining, технологий обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Исходя из выше сказанного, можно утверждать, что да data mining – это ИАД, следовательно, такой подход будет уместен в таких областях как прикладная статистика, искусственный интеллект, визуализация, распознавание образов и т.п. (рис. 1)



Рисунок 1 – Data Mining как мультидисциплинарная область

В связи с тем, что методы Data Mining постоянно развиваются, единого мнения какие задачи следует относить к данной технологии нет. Большинство источников выделяют предсказательные и описательные задачи (рис. 2).

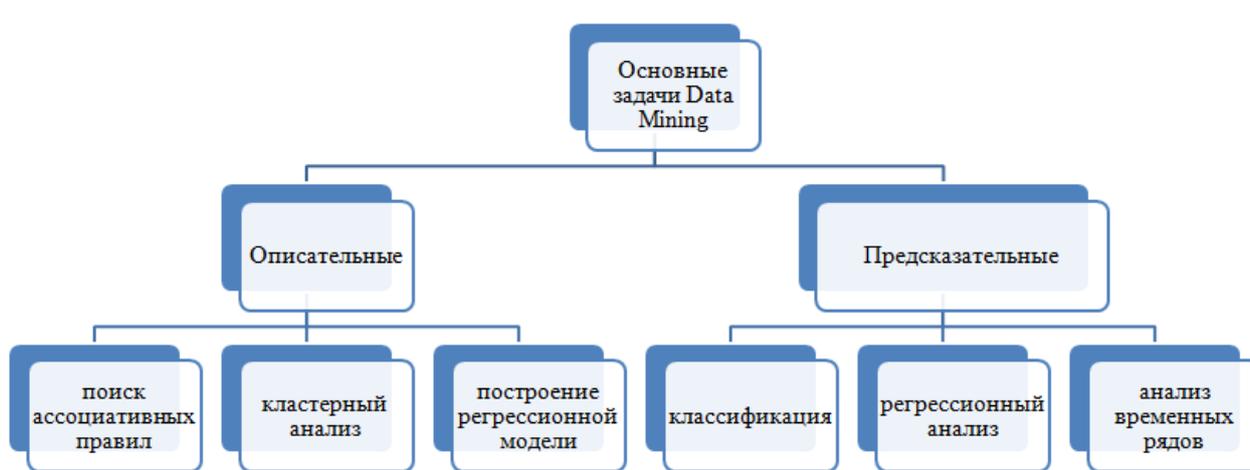


Рисунок 2 – Основные задачи Data Mining

В настоящее время можно выделить два направления, в которых используют технологию Data Mining – это массовый продукт и инструмент научных исследований. Основные сферы, в которых реализуется данная технология, представлены на рисунке 3.

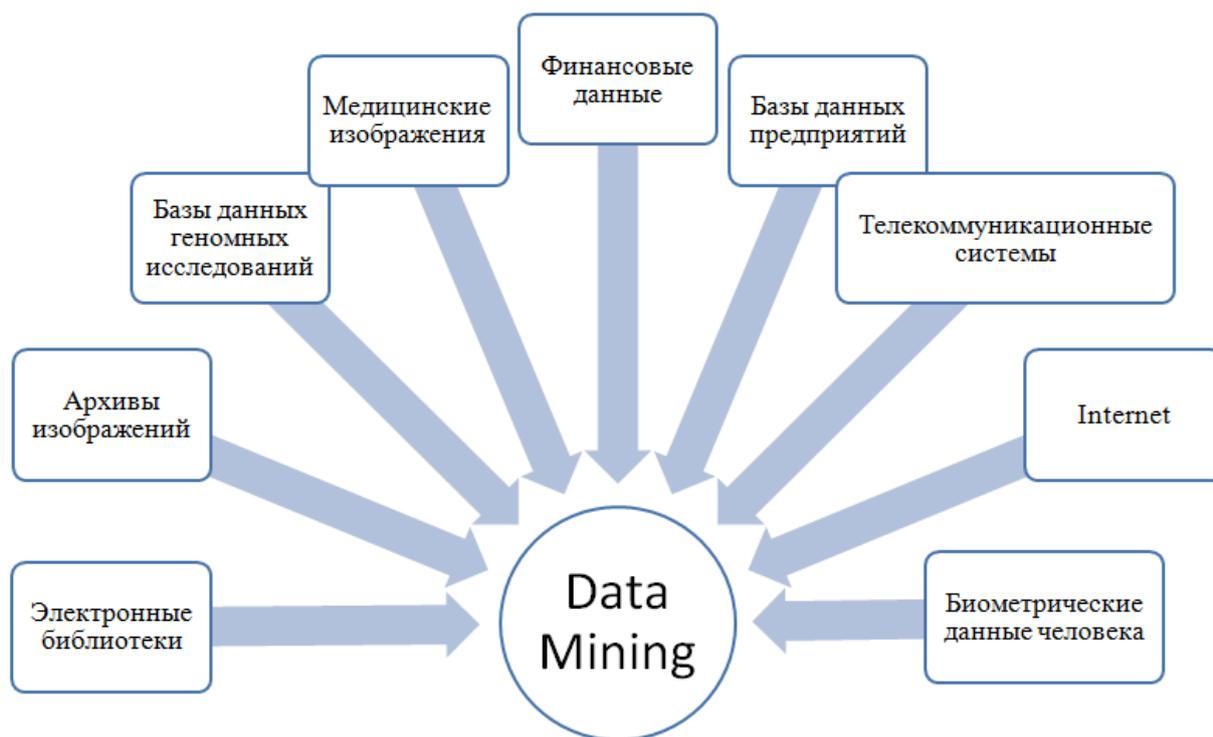


Рисунок 3 – Сферы применения технологии Data Mining

Как видно из данных (рис. 3) технология Data Mining используется практически во всех сферах деятельности человека, где накоплены ретроспективные данные.

**Глава 2.** Процесс использования технологий Data mining включает в себя несколько этапов, традиционным считают следующее деление: 1) анализ предметной области; 2) постановка задачи; 3) подготовка данных; 4) построение моделей; 5) проверка и оценка моделей; 6) применение модели; 7) коррекция и обновление модели (рис. 4).

*Этап 1.* При анализе области исследований необходимо создание модели, которая будет описывать/объяснять все процессы, происходящие в ней, а также какие данные используются и как они взаимосвязаны друг с другом. Чем детальнее будет описана область, тем лучше полученный результат. Предметную область, как правило, описывают с помощью методик структурного анализа SADT или IDEF0, объектно-ориентированного анализа UML и т.п.

*Этап 2.* На данном этапе необходимо четко определить и сформулировать задачи исследования. Здесь же необходимо рассмотреть способы учёта результатов моделирования. Задачи исследования должны отражать следующие моменты: что и как искать, какие типы связей, необходимы прогнозы или нет, что будем иметь в качестве результата, требуется ли очистка или обработка данных, какие статистические закономерности используются и т.п.

Следует помнить, что технологии интеллектуального анализа данных не в состоянии заменить аналитиком и ответить на вопросы, которые не были сформулированы, поэтому на данном этапе необходимо учесть большое количество возможных ситуаций и попытаться их описать.

*Этап 3.* Самый ответственный шаг, т.к. именно на данном этапе создается база данных для Data Mining и, следовательно, от этого этапа зависит качество полученного результата анализа. Данный этап можно условно поделить на несколько этапов поменьше.

*Этап 3.1. Определение и анализ требований к данным.* Как понятно из названия, здесь необходимо смоделировать данные для исследования, в связи с чем, изучаются вопросы доступа к данным, наличия или отсутствия внешних/внутренних источников этих данных, возможное их размещение, функциональную и организационную составляющую и многое другое.

*Этап 3.2. Сбор данных.* На данном шаге требуется определить источник исследуемых данных – это может быть собственное хранилище данных, либо базы данных существующих информационных систем (архивные, справочные, оперативные, пользовательские).

При решении конкретных задач для анализа используются не все данные из базы, в связи с чем, задаются необходимые критерии к выбору требуемого подмножества данных, и только оно подвергается анализу. Стоит отметить, что при выборе данных следует описывать максимально возможное количество факторов и признаков, и не делать упор на показатели процесса. Кроме того выбранные данные могут быть упорядоченными или неупорядоченными, что накладывает дополнительные условия на созданную интеллектуальную модель.

*Этап 3.3. Предварительная обработка данных.* Основная цель данного шага заключается в оценивании качества данных выбранных на этапе 3.2. Следовательно, все данные классифицируются как данные высокого и низкого качества.

Данные высокого качества – это полные, точные, своевременные данные, которые поддаются интерпретации.

Данные низкого качества – это отсутствующие, неточные или бесполезные данные с точки зрения практического применения, среди них выделяют пропущенные значения, дубликаты, шумы и выбросы.

Как видно из рис. 4, на данном шаге может потребоваться очистка или обогащение данных в связи с наличием или отсутствием данных высокого или низкого качества. Более подробно про методы и инструменты очистки данных описано в работах [115-118]. В зависимости от детализации данных, они подлежат кодированию, что облегчает использование алгоритмов распознавания образов. Способ кодирования информации также будет определять конечный результат.

*Этапы 4 – 6.* Знания, полученные на предыдущих этапах, определяют тип модели, которая будет использоваться для анализа. Для создания модели применяют всевозможные алгоритмы/методы интеллектуального анализа данных, как самостоятельно, так и в комбинации, что позволяет создать модель наиболее приближенную к описываемому объекту или процессу. (рис. 5). Таким образом, выбор метода моделирования напрямую зависит от постановки задачи, ее специфики и результатов.

Далее оценивается определение точность модели и эффективность выбранного алгоритма.

*Этап 7.* По мере использования модели исходные данные могут обновляться, следовательно, необходимо периодически её корректировать, для этого модель обучается заново, в противном случае, она потеряет свою функциональность и адекват-

ность. Кроме того сам процесс создания и обучения модели сопровождается накоплением погрешности, т.к. создать идеальную модель невозможно. В случае технологии Data Mining наиболее распространенными погрешностями модели являются: неверные/недостовверные исходные допущения; ограниченные возможности на этапе сбора данных; неуверенность пользователя; завышенная стоимость проекта.



Рисунок 4 – Процесс обработки данных

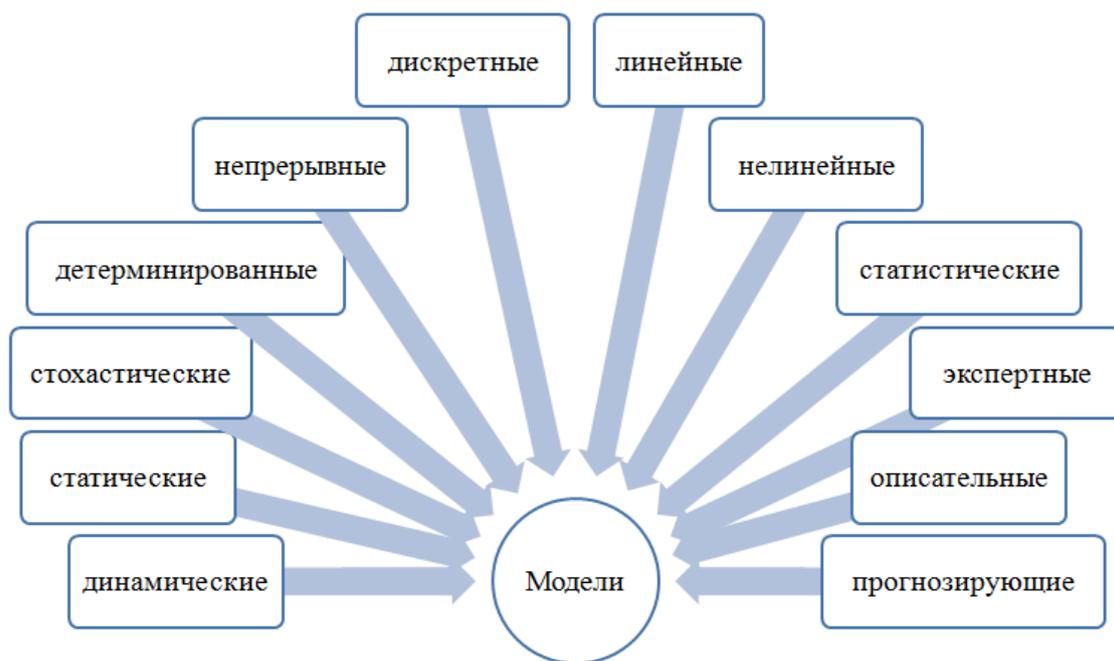


Рисунок 5 – Классификация моделей

**В главе 3** названы инструменты Data Mining и рассматриваются возможности для анализа данных пакета Excel на примере задачи «Дифракция света на отверстии».

В настоящее время инструменты интеллектуального анализа данных хорошо представлены на рынке [119-120].

- Согласно данным KDnuggets [121] все инструменты можно поделить на 4 класса: общего и специального назначения, коммерческие и бесплатные. Каждый из классов инструментов Data Mining определяется по следующим категориям:
- наборы инструментов:
  - классификация данных;
  - кластеризация и сегментация;
  - инструменты статистического анализа;
  - анализ текстов, извлечение отклонений;
  - инструменты визуализации.
- В категорию набор инструментов входят универсальные инструменты, применяемые для решения задач классификации и кластеризации, например [122]:
- Clementine (<http://www.spss.com/clementine>) – бизнес-процесс для снижения времени принятия решений;
- DBMiner 2.0 Enterprise (<http://www.dbminer.com>) – для исследования Big Data;
- IBM Intelligent Miner for Data (<http://www.ibm.com/software/data/iminer/fordata/>) – поддерживает полный Data Mining процесс;
- Oracle Data Mining (ODM) (<http://otn.oracle.com/products/bi/9idmining.html>) – обеспечивает GUI, PL/SQL-интерфейсы, Java-интерфейс;
- Polyanalyst (<http://www.megaputer.com/>) – обеспечивающий всесторонний Data Mining;
- SAS Enterprise Miner (<http://www.sas.com/>) – интегрированный набор, который обеспечивает дружелюбный GUI. Поддерживается методология SEMMA.
- SPSS (<http://www.spss.com/clementine/>) – обеспечивающий всесторонний Data Mining;
- Statistica Data Miner (<http://www.StatSoft.com/>) – обеспечивающий всесторонний Data Mining, а также приложение разработки систем;
- Loginom (<https://loginom.ru/>) – приложение, выполняющая импорт, обработку, визуализацию и экспорт данных в форме OLAP кубов, отчетов, моделей и закономерностей;
- Weka (<http://www.cs.waikato.ac.nz/ml/weka/index.html>) – бесплатный набор алгоритмов машинного обучения для решения реальных Data Mining-проблем.
- В следующих категориях инструменты интеллектуального анализа данных способны решать задачи с помощью различных статистических и кибернетических методов:
- Azmy SuperQuery (<http://www.azmy.com/>), поисковик ассоциативных правил;
- FIMI, Frequent Itemset Mining Implementations (<http://fimi.cs.helsinki.fi/>) – является репозиторием, включающим программное обеспечение и БД;

- ClustanGraphics3, (<http://www.clustan.com/>) – иерархический кластерный анализ "сверху вниз", поддерживаются мощные графические возможности;
- Visipoint (<http://www.visipoint.fi/>) – кластеризация методом самоорганизующихся Карт Кохонена и визуализация;
- MCLUST/EMCLUST ([http://www.stat.washington.edu/fraley/mclust\\_home.html](http://www.stat.washington.edu/fraley/mclust_home.html)) – создание кластеров при помощи модельного подхода и дискриминантного анализа, иерархическая кластеризация;
- ReCkless (<http://cde.iiit.net/RNNs/>) – кластерный алгоритм, основанный на концепции k-ближайших соседей, способен проводить поиск и идентификацию шумов и выбросов для уменьшения их влияния на результаты кластеризации;
- ExcelNeuralPackage ([http://www.neurok.ru/demo/enp/demo\\_enp.htm](http://www.neurok.ru/demo/enp/demo_enp.htm)) – Excel-надстройки, предназначены для решения задач прогнозирования и оценивания с использованием нейронных сетей.

## ЗАКЛЮЧЕНИЕ

В выпускной квалификационной работе дается принципиальное представление об интеллектуальном анализе данных, рассматриваются основные методы создания моделей и инструменты для их анализа.

Проверена работоспособность приложения MS Excel «Анализ данных». На экспериментальных данных проведен статистический и регрессионный анализ.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Cox M., Ellsworth D. Application-Controlled Demand Paging for Out-of-Core Visualization // Proceedings of the 8th IEEE Visualization '97 Conference. – 1997. – Vol. 1070-2385. – P. 235-244. URL: [paging\\_outofcore\\_viz97.pdf](http://paging_outofcore_viz97.pdf) (uic.edu) (Дата обращения 30.03.2021 г.). – Загл. с экрана. – Яз. англ.
2. Press G. A Very Short History of Big Data [Электронный ресурс] // Forbes.com. 09.05.2013. URL: <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-shorthistory-of-big-data/#92208b965a18> (Дата обращения: 30.03.2021). – Загл. с экрана. – Яз. англ.
3. Корнев М. С. История понятия "большие данные" (Big Data): словари, научная и деловая периодика // Вестник РГГУ. Серия: Литературоведение. Языкознание. Культурология. – 2018. – №1 (34). – С. 81-85.
4. Lynch C. Big data: how do your data grow? // Nature. – 2008. – Vol. 455. – № 7209. – P. 28–29. DOI: 10.1038/455028a
5. McKinsey Global Institute. Big data: The next frontier for innovation, competition, and productivity. [Электронный ресурс] 2011. – 146 p. URL: [MGI\\_big\\_data\\_full\\_report.aspx](http://mgi.com/mgi_big_data_full_report.aspx) (mckinsey.com) (Дата обращения 30.03.2021 г.). – Загл. с экрана. – Яз. англ.
6. Роднин А.В., Турчик В.Ю. Концепция применения интеллектуального анализа данных в средствах защиты информации базы данных // Физика. Технология. Инновации: Сб.науч. тр. – 2015. – с. 263-269. URL: [https://elar.urfu.ru/bitstream/10995/35647/1/fti\\_2015\\_43.pdf](https://elar.urfu.ru/bitstream/10995/35647/1/fti_2015_43.pdf) (Дата обращения 30.03.2021 г.). – Загл. с экрана. – Яз. рус.
7. Левкович-Маслюк Л. Великие раскопки и великие вызовы // Компьютерра. – 2007. – № 11 (679). – С. 48-51.
8. Слесарев Е. В., Тесля В. В. Проблемы, тенденции развития и перспективные направления. [Электронный ресурс] // Электроника и информационные технологии. – 2011. – Вып. 2 (11). URL:

- [http://fetmag.mrsu.ru/2011-2/pdf/applications\\_of\\_data\\_mining.pdf](http://fetmag.mrsu.ru/2011-2/pdf/applications_of_data_mining.pdf) (Дата обращения 20.03.2021 г.). – Загл. с экрана. – Яз. рус.
9. Савченко Л.М., Бежитский С.С. DataMining и области его применения // Актуальные проблемы авиации и космонавтики. – 2015. – №11. – С. 611-613.
  10. Дюк В. А. Data Mining – интеллектуальный анализ данных [Электронный ресурс] // iTeam [сайт] URL: <https://blog.iteam.ru/data-mining-intellektualnyj-analiz-dannyh/> (Дата обращения 10.03.2021). – Загл. с экрана. – Яз. рус.
  11. Преображенский Ю.П., Мясников О.А. Анализ перспектив информационных технологий в сфере интернет вещей // Вестник Воронежского института высоких технологий. – 2020. – № 1 (32). – С. 43-45.
  12. Mukhopadhyay A. A. Survey of Multiobjective Evolutionary Algorithms for Data Mining / Mukhopadhyay A., Maulik U., Bandyopadhyay S., Coello C.A. IEEE Transactions on Evolutionary Computation. – 2014. – V.18. – N1. – P. 4-35.
  13. Полковникова Н.А., Курейчик В.М. Об интеллектуальном анализе баз данных для экспертной системы // Информатика, вычислительная техника и инженерное образование. – 2013. – № 2 (13). С. 46-59.
  14. Bathla G., Aggarwal H., Rani R. Migrating From Data Mining to Big Data Mining. // International Journal of Engineering & Technology, [S.l.]. –2018. – Vol. 7. N. 3.4. – P. 13-16. DOI: 10.14419/ijet.v7i3.4.14667.
  15. Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining при медико-лабораторных исследованиях для выявления закономерностей в массивах данных // Врач и информационные технологии. – 2005. – № 2. – С. 49–56.
  16. Замятин А.В., Марков Н.Г., Напряшкин А.А. Адаптивный алгоритм классификации с использованием текстурного анализа для автоматизированной интерпретации аэрокосмических изображений // Исследование Земли из космоса. – 2004. – № 2. – С. 32–40.
  17. Чубукова И. А. Data Mining. Учебное пособие. – М.: Интернет-университет информационных технологий - ИНТУИТ.ру, БИНОМ. Лаборатория знаний, 2008. – 384 с.
  18. Дядичев В. В., Ромашка Е. В., Голуб Т. В. Задачи и методы интеллектуального анализа данных // Геополитика и экогеодинамика регионов. 2015. – Том 1 (11). – Вып. 3. – С. 23–29.
  19. Ian H. Witten Data Mining: Practical Machine Learning Tools and Techniques / Ian H. Witten, Eibe Frank and Mark A. Hall. ; 3rd Edition. – Morgan Kaufmann, 2011. – 664 p.
  20. Чернышова Г.Ю. Интеллектуальный анализ данных: учеб. пособие. –Саратов: Изд-во СГСЭУ, 2012. – 92 с.
  21. Макарычев П. П., Афонин А. Ю. Оперативный и интеллектуальный анализ данных: учеб. пособие. – Пенза: Изд-во ПГУ, 2010. – 156 с.
  22. Тучкова А.С., Кондрашева П.П. Термин «Data mining». Задачи, решаемые методами data mining // Тенденции развития науки и образования. – 2019. – №55-2. – С.27-30. DOI: 10.18411/ij-10-2019-26
  23. Шумейко А. А. Интеллектуальный анализ данных (Введение в Data Mining): учеб. пособ. / А. А. Шумейко, С. Л. Сотник. – Днепропетровск: Белая Е. А., 2012. – 212 с.
  24. Нестеров С. А. Интеллектуальный анализ данных средствами MS SQL Server 2008. – Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. – 303 с.
  25. Петрунин Ю. Ю. Информационные технологии анализа данных. Data Analysis. – М.: КДУ, 2010. – 292 с.
  26. Горлушкина Н. Н., Коцюба И. Ю., Хлопотов М. В. Задачи и методы интеллектуального анализа образов // ОТО. – 2015. – №1. URL: <https://cyberleninka.ru/article/n/zadachi-i-metody-intellektualnogo-analiza-obrazova> (Дата обращения: 23.03.2021). – Загл. с экрана. – Яз. рус.
  27. Салмин А.А., Душенко О.Д., Мурын И.С., Полянчинков В.А. Задачи, решаемые методами Data Mining // Актуальные проблемы современной науки. – 2020. – №5 (114). – С. 174-175.
  28. Mehmed Kantardzic Data Mining: Concepts, Models, Methods, and Algorithms. – Wiley-IEEE Press, 2019. – 672 p. DOI: 10.1002/9781119516057.

29. Busarov V., Grafeeva N., Mikhailova E. (2016) A Comparative Analysis of Algorithms for Mining Frequent Itemsets. In: Arnicans G., Arnican V., Borzovs J., Niedrite L. (eds) Databases and Information Systems. DB&IS 2016. Communications in Computer and Information Science, vol 615. Springer, Cham. DOI: 10.1007/978-3-319-40180-5\_10
30. Степанов П.Г. Технология Data Mining: Интеллектуальный анализ данных. – Казань: Изд-во КГУ, 2008. – 57 с.
31. Sel A. Veri madenciligi kümeleme yöntemleri kullanarak karbon emisyonu göstergeleri açısından oecd ülkelerinin siniflandırılması // Karadeniz Uluslararası Bilimsel Dergi. – 2020. – Vol. 46. – P. 169-187. DOI: 10.17498/kdeniz.679555
32. Заболотный А.С. Информационная подсистема интеллектуального анализа данных на основе методов Data Mining // Научный альманах. – 2016. – № 5-3(19). – С. 75-77.
33. DataMining - технология добычи данных [Электронный ресурс]. URL: <http://bourabai.ru/tpoi/datamining.htm> (Дата обращения: 27.03.2021). – Загл. с экрана. – Яз. рус.
34. Ломакин Н.И., Московцев А.Ф., Копылов А.В., Телятникова В.С., Самородова И.А., Горбунова А. В., Попова Я.А., Хабарова А.Д., Кива Ю.Ю., Максимова О.Н. Data mining с искусственной нейронной сетью для определения инвестиционной привлекательности регионов // Наука Красноярья. – 2017. – Т. 6. – № 4-2. – С. 98-102.
35. Костюкова Н.И. Применение технологии Data Mining для решения задач оптимизации проектирования сложных технических систем // Альманах современной науки и образования. Тамбов : Грамота. – 2010. – № 5 (36). – С. 60–61.
36. Марухина О.В., Мокина Е.Е., Берестнева Е.В. Применение методов data mining для выявления скрытых закономерностей в задачах анализа медицинских данных // Фундаментальные исследования. – 2015. – № 4. – С. 107-113.
37. Пальмов С. В. Использование нейронных сетей в стоматологии / С. В. Пальмов, А. В. Бахмурина // Проблемы развития предприятий: теория и практика. – 2020. – № 1-2. – С. 237-240.
38. Абруков В., Абруков С., Смирнов А., Карлович Е. (2013). Data mining в научных исследованиях. DOI: 10.13140/2.1.4664.0962.
39. Назаров Л.Е. Применение многослойных нейронных сетей для классификации земных объектов на основе анализа многозональных сканерных изображений // Исследование Земли из космоса. – 2000. – № 6. – С. 41–50.
40. Герасимов С.В., Мещеряков А.В., Колосов И.Ю., Глотов Е.С., Попов И.С. Обработка больших объемов сырых астрономических данных с помощью модели вычислений MapReduce // Труды ИСП РАН. – 2015. – №6. – С. 315-322
41. Ivezić Z., Connolly A., VanderPlas J., Gray A. Statistics, Data Mining, and Machine Learning in Astronomy. – Princeton University Press, 2013. – 552 p.
42. Пальмов, С. В. Интеллектуальный анализ текста / С. В. Пальмов, А. С. Мячина // Евразийское Научное Объединение. – 2020. – № 11-2(69). – С. 116-122. – DOI 10.5281/zenodo.4313358.
43. Полковникова Н.А. Проектирование гибридной экспертной системы поддержки принятия решений // В сб. тез. докл. II Всероссийского конгресса молодых учёных. – СПб: НИУ ИТМО, 2013 – Вып. I. – С. 46-48.
44. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям / Н.Б.Паклин, В.И. Орешков. – СПб: Питер, 2013. – 704 с.
45. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб: БХВ-Петербург, 2004. – 336 с.
46. Амаева Л.А. Сравнительный анализ методов интеллектуального анализа данных // Инновационная наука. – 2017. – №2-1. – С.27-29.
47. Вентцель, Е. С. Теория вероятностей и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров. – М.: Наука, 1988. – 416 с.
48. Хлевнюк А. Основы линейной регрессии [Электронный ресурс] / Хабр [сайт]. URL: <https://habr.com/ru/post/514818/> (Дата обращения 03.04.2021 г.) – Загл. с экрана. – Яз. рус.,
49. Saraswat M. Beginners Guide to Regression Analysis and Plot Interpretations [Электронный ресурс] / Hacherearth [сайт]. URL: <https://www.hackerearth.com/practice/machine-learning/machine->

- learning-algorithms/beginners-guide-regression-analysis-plot-interpretations/tutorial/ (Дата обращения 03.04.2021 г.) – Загл. с экрана. – Яз. англ.,
50. Кобзарь А. И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 816 с.
51. Григорян М. R – значит регрессия [Электронный ресурс] / Хабр [сайт]. URL: <https://habr.com/ru/post/350668/> (Дата обращения 03.04.2021 г.) – Загл. с экрана. – Яз. рус.
52. Skhvediani A. Основные предпосылки классической линейной регрессии и последствия их нарушений // Инновации и инвестиции. – 2020. – №8. – С. 38-42.
53. Демин С. Е., Демина Е. Л. Математическая статистика: учеб.-метод. пособие. – Нижний Тагил: НТИ (филиал) УрФУ, 2016. – 284 с.
54. Шахиди А. Деревья решений: общие принципы [Электронный ресурс] / Loginom [сайт] URL: <https://loginom.ru/blog/decision-tree-p1> (Дата обращения 03.04.2021 г.) – Загл. с экрана. – Яз. рус.
55. Груздев А. Деревья решений [Электронный ресурс]. URL: <https://dmkpress.com/files/PDF/978-5-97060-456-4.pdf>. (Дата обращения 03.04.2021 г.) – Загл. с экрана. – Яз. рус.
56. Митрофанов С.А. О модификации алгоритма обучения дерева решений // Решетневские чтения. – 2018. – Т. 2. – С. 138-139.
57. Breiman L., Friedman J., Olshen R., Stone C. Classification And Regression Trees. [Электронный ресурс]. – Wadsworth, Belmont, California: Boca Raton, 2017. – 368 p. DOI: 10.1201/9781315139470.
58. Quinlan J.R. Induction of decision trees // Kluwer Academic Publishers. 1986. – Vol.1(1). – P. 81-106.,
59. Quinlan R. C4.5: Programs for Machine Learning. – San Mateo, CA: Morgan Kaufmann, 1993. – 302 p.
60. Generation of User Interest Ontology Using ID3 Algorithm in the Social Web / Jong-Soo Sohn, Qing Wang, In-Jeong Chung // IT Convergence and Security 2012. Lecture Notes in Electrical Engineering, 2012. – P. 1067-1074.
61. Salzberg S., Segre A. Review of C4.5: Programs for Machine Learning by J. Ross Quinlan // Machine Learning, 1994. – Vol. 16. P. 235-240. DOI: 10.1007/BF00993309.
62. Некрасов М.В. Автоматизация метода «дерево решений» // Актуальные вопросы экономических наук. – 2013. – №32. – С. 66-70.
63. Андреев И. Деревья решений – CART математический аппарат. [Электронный ресурс]. / BaseGroup Labs [сайт]. URL [http://www.basegroup.ru/library/analysis/tree/math\\_cart\\_part1/](http://www.basegroup.ru/library/analysis/tree/math_cart_part1/) (Дата обращения: 03.04.2021) – Загл. с экрана. – Яз. рус.
64. Singla A. Real-time swing-up and stabilization control of a cart-pendulum system with constrained cart movement / A. Singla, G. Singh // International Journal of Nonlinear Sciences and Numerical Simulation. – 2017. – Vol. 18. – No 6. – P. 1-15. DOI 10.1515/ijnsns-2017-0040.
65. Miftakhova A. A. Presenting classification opportunities of cart algorithm // Science and World. – 2017. – Vol. 1. – № 4(44). – P. 70-72.
66. Breiman L., Friedman J., Olshen R., Stone C. Classification And Regression Trees. – New York: Chapman and Hall, Wadsworth, 1983. – 368 P.
67. Дюк В. А. Современные технологии обнаружения знаний в базах данных / В. А. Дюк, Л. В. Калягина // Вестник КрасГАУ. – 2004. – № 4. – С. 27-32.
68. Воронцов К. В. Лекции по методу опорных векторов [Электронный ресурс]. 2007. – 18 с. URL: <http://www.ccas.ru/voron/download/SVM.pdf>. (Дата обращения: 03.04.2021) – Загл. с экрана. – Яз. рус.
69. Китова О. В., Колмаков И. Б., Пеньков И. А. Метод машин опорных векторов для прогнозирования показателей инвестиций // Статистика и экономика. – 2016. – №4. – С. 27-30.
70. Дмитриев Е. А. Метод опорных векторов // Научные исследования и разработки студентов: Сб. матер. IV Межд. студ. научно-практической конф., Чебоксары, 29 июня 2017 года / Редколлегия: О.Н. Широков [и др.]. – Чебоксары: ООО "Центр научного сотрудничества "Интерактив плюс", 2017. – С. 131-132.

71. Мальчиц В. С. Обработка данных для машинного обучения и применение метода опорных векторов для реализации классификатора новостей / В. С. Мальчиц, А. Н. Гетман // Вестник Амурского государственного университета. Серия: Естественные и экономические науки. – 2019. – № 87. – С. 8-13. DOI 10.22250/jasu.2019.87.8-13.
72. Пулатов Б. Н. Метод ближайших соседей в задачах непараметрического оценивания регрессии: специальность 01.01.05 "Теория вероятностей и математическая статистика": автореферат дисс. на соиск. уч. ст. канд. физ.-мат. наук / Пулатов Бахрам Нигматович. – Ташкент, 1993. – 16 с.
73. Тярт, Н. А. Метод к ближайших соседей и его программная реализация для автоматизированной классификации объектов / Н. А. Тярт, Н. Н. Степанов // Современные технологии в науке и образовании – СТНО-2016: сб. тр. межд. научно-техн. и научно-метод. конф.: в 4 томах, Рязань, 02–04 марта 2016 года / Под общ. редакцией О.В. Миловзорова. – Рязань: РГРУ, 2016. – С. 7-10.
74. Li Q. B. Method for fine pattern recognition of space targets using the entropy weight fuzzy-rough nearest neighbor algorithm / Q. B. Li, Y. Wei, W. J. Li // Zurnal prikladnoj spektroskopii. – 2020. – Vol. 87. – No 6. – P. 886-890.
75. Хлопотов М. В. Применение байесовской сети при построении моделей для оценки уровня сформированности компетенций // Вестник евразийской науки. Интернет-журнал. – 2014. – №5 (24). – С. 1-28.
76. Бидюк П.И., Терентьев А.Н. Построение и методы обучения байесовских сетей // ТВИМ. – 2004. – №2. – С. 139-154.
77. Арустамов С. А., Дайнеко В. Ю. Применение динамической байесовской сети в системах обнаружения вторжений // Научно-технический вестник информационных технологий, механики и оптики. – 2012. – №3 (79). – С. 128-133.
78. Heckerman, D. A tutorial on learning with Bayesian networks // Learning in graphical models // MIT Press, 1998. – P. 301–354.
79. Згуровский М.З., Бидюк П.И. Методы построения байесовских сетей на основе оценочных функций // Кибернетика и системный анализ. – 2008. – № 2. – С. 81-88.
80. Daly R., Shen Q., Aitken S. Learning Bayesian networks: Approaches and issues. // The Knowledge Engineering Review. – 2011. – Vol. 26(2). – P. 99-157. DOI:10.1017/S0269888910000251.
81. Scutari M. Learning bayesian networks with the bnlearn package // Journal of Statistical Software. – 2010. – № 35. – P. 1–22.
82. Азарнова Т. В., Баркалов С. А., Полухин П. В. Разработка гибридного алгоритма обучения структуры динамической байесовской сети на основе метода Левенберга - Марквардта // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. – 2018. – №4. – С. 16-24.
83. Mittal A., Kassim, A. (2007). Bayesian Network Technologies: Applications and Graphical Models. – Hershey, PA: IGI Global, 2007. – 368 p. DOI: 10.4018/978-1-59904-141-4.
84. Scutari M., Millan E. A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation // User Modeling and UserAdapted Interaction. – 2002. – № 12. – P. 281–330.
85. Макарычев П. П., Афонин А. Ю. Оперативный и интеллектуальный анализ данных: учеб. пособие. – Пенза: Изд-во ПГУ, 2010. – 156 с.
86. Зубков Е. В., Белов В. М. Методы интеллектуального анализа данных и обнаружение вторжений // Вестник СибГУТИ. – 2016. - № 1. – С. 118-133.
87. Гафаров Ф.М. Искусственные нейронные сети и приложения: учеб. пособие / Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018. – 121 с.
88. Слепенков В. О. Использование технологий машинного обучения для создания искусственных нейронных сетей / В. О. Слепенков, А. В. Бондарь // Молодежный научно-технический вестник. – 2016. – № 4. – С. 20.
89. Розенблат Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга. – М.: Мир, 1965. – 481 с.
90. Манжула В.Г., Федяшов Д.С. Нейронные сети кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Фундаментальные исследования. – 2011. – № 4. – С. 108-115.

91. Кутыркин А.В., Сёмин А.В. Использование нейронной сети Хопфилда для решения оптимизационных задач маршрутизации: Методические указания. – М.: МИИТ, 2007. – 15 с.
92. Павлова А. И., Бобкова К. А. Сравнение алгоритмов распознавания образов нейронными сетями хопфилда // В мире научных открытий. – 2016. – №5 (77). – С. 134-145. DOI: 10.12731/wsd-2016-5-7.
93. Переверзев И. А. Эффективные алгоритмы построения нейронной сети на основании оценки входных параметров (глубокое машинное обучение) / И. А. Переверзев, В. Н. Квасницкий // Вестник Московского финансово-юридического университета. – 2016. – № 1. – С. 253-261.
94. Мамхягов А. З. Нейронные сети: машинное обучение, искусственный интеллект - зачем нам думать // Молодежь и наука: реальность и будущее : Материалы X межд. научно-практ. конф., Невинномысск, 01 марта 2017 года / Редколлегия: Т.Н. Рябченко, Е.И. Бурьянова. – Невинномысск: НОУ ВПО "Невинномысский институт экономики, управления и права", 2017. – С. 93-96.
95. Тихонов А. А. Большие данные и глубокое машинное обучение в искусственных нейронных сетях // Наука и образование сегодня. – 2018. – № 6(29). – С. 35-38.
96. Zheng S., Jayasumana S., Romera-Paredes B., Vineet V., Su Z., Du D., Huang C., Torr P. H. S., Conditional Random Fields as Recurrent Neural Networks // Proceedings of the 2015 IEEE International Conference on Computer Vision. – 2015. – P. 1529–1537.
97. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P. 770–778.
98. Rashchenko Y. V., Kozynchenko V. A. New algorithms of an artificial neural network ART-1 training // IEEE 2015 International Conference «Stability and Control Processes» in memory of V. I. Zubov (SCP). – 2015. – P. 663–664.
99. Глубокое обучение для новичков: распознаем изображения с помощью сверточных сетей / Хабр [Сайт] URL: <https://habrahabr.ru/company/wunderfund/blog/314872/> (Дата обращения: 03.04.2021) – Загл. с экрана. – Яз. рус.
100. Белоглазов Д. А. Особенности нейросетевых решений, достоинства и недостатки, перспективы применения // Известия ЮФУ. Технические науки. – 2008. – №7. – С. 105-110.
101. Махрусе Насма Современные тенденции методов интеллектуального анализа данных: метод кластеризации // Московский экономический журнал. – 2019. – №6. URL: <https://cyberleninka.ru/article/n/sovremennye-tendentsii-metodov-intellektualnogo-analiza-dannyh-metod-klasterizatsii> (Дата обращения: 03.04.2021) – Загл. с экрана. – Яз. рус.
102. Демин Г. А. Методы принятия управленческих решений [Электронный ресурс]: учеб. пособие. – Пермь: Перм. гос. нац. исслед. ун-т, 2019. – 88 с. – URL: <http://www.psu.ru/files/docs/science/books/uchebnie-posobiya/demin-metody-prinyatiya-upravlencheskikh-reshenij.pdf> (Дата обращения: 03.04.2021) – Загл. с экрана. – Яз. рус.
103. Хорольцева Е.Б., Федорова А.В. Исследование систем управления: методологический аспект // Вестник Поволжского института управления. – 2013. – № 2 (35). – С. 110–117.
104. Фаломкина, О. В. Исследование нечетких и неопределенных нечетких методов анализа и интерпретации данных: специальность 05.13.18 "Математическое моделирование, численные методы и комплексы программ": дисс. на соискание уч. ст. канд. физ.-мат. наук / Фаломкина Олеся Владимировна. – Москва, 2006. – 141 с.
105. Зиновьев А. Ю. Визуализация многомерных данных. – Красноярск: Изд-во КГТУ, 2000. — 180 с.
106. Математические методы анализа и интерпретация социологических данных / В.Г. Андреенков, К.Д. Аргунова, В.И. Паниотто и др./ Отв. ред.: В.Г. Андреенков, Ю.Н. Толстова; АН СССР, Ин-т социологии. – М.:Наука,1989. – 175 с.
107. Зиновьев А. Ю., Носс И. Н. Методы интерпретации эмпирических данных в психологическом исследовании, рефракции интерпретации и их контроль // Человеческий капитал. – 2014. – № 7(67). – С. 16-22.
108. Романова И.К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах // Машиностроение и компьютерные технологии. – 2016. – №3. – С. 133-167..

109. Гнеденко Б.В., Беляев Ю.К., Соловьев А.Д. Математические методы в теории надежности: Основные характеристики надежности и их статистический анализ. Изд.2. – М.: URSS, 2013. – 584 с.
110. Пупков К. А. Применение интеллектуальных технологий - перспективное направление развития теории и практики проектирования и создания систем обработки информации и управления // Вестник РУДН. Серия: Инженерные исследования. – 2008. – №4. – С. 44-52.
111. Интеллектуальные информационные системы и технологии: Монография / А.В. Остроух, Н.Е. Суркова. – Красноярск: Научно-инновационный центр, 2015. – 370 с.
112. Кучковская Н. В. Экономическое значение систем интеллектуального управления предприятием // Инновационное развитие экономики. – 2019. – № 1(49). – С. 44-50.
113. Коберн А. Современные методы описания функциональных требований к системам. : Пер. с англ. – М.: Лори, 2002. – 263 с.
114. Методология функционального проектирования IDEF0: Учеб. пособие / В.В. Бахтизин, Л.А. Глухова. – Мн.: БГУИР, 2003. – 24 с.
115. Рам Э., Хонг Х. Очистка данных: проблемы и актуальные подходы [Электронный ресурс] / Журнал BLM World, 2000. – URL: <http://www.iso.ru/rus /document5815.phtml>. (Дата обращения: 04.04.2021) – Загл. с экрана. – Яз. рус.
116. Луконькин А. Варианты очистки базы данных // Системный администратор. – 2008. – № 12(73). – С. 55.
117. Обзор способов очистки данных на примере Google Refine / Ю. В. Катков, А. Н. Злобин, А. Алексишин, И. А. Мамонтов // Управление знаниями и технологии семантического веба. – 2010. – № 1. – С. 175-178.
118. Ларина Ю. А. Основы объектно-ориентированного моделирования с использованием языка UML: учебное пособие. Ярославль: ЯрГУ, 2010. – 151 с.
119. 25 лучших инструментов для интеллектуального анализа данных [Электронный ресурс] / CoderLessons.com [Сайт] URL: <https://coderlessons .com/tutorials/bolshie-dannye-i-analitika/teoriia-khraneniia-dannykh/31-25-luchshikh-instrumentov-dlia-intellektualnogo-analiza-dannykh> (Дата обращения: 04.04.2021) – Загл. с экрана. – Яз. рус.
120. Российская VI-платформа [Электронный ресурс] / Триафлай [Сайт] URL: <https://triafly.ru/> (Дата обращения: 04.04.2021) – Загл. с экрана. – Яз. рус.
121. KDnuggets [Сайт] URL: <https://www.kdnuggets.com/> (Дата обращения: 04.04.2021) – Загл. с экрана. – Яз. англ.
122. Инструменты Data Mining / ЦРКБИ [Сайт] URL: <https://hsbi.hse.ru/articles/instrumenty-data-mining/> (Дата обращения: 04.04.2021) – Загл. с экрана. – Яз. рус.