ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS

УДК 004. 82: 004.89
DOI: 10.21822/2073-6185-2023-50-4-121-128 Оригинальная статья /Original article

CC BY 4.0

Модель интеллектуального анализа данных для исследования системы квот вылова водных биологических ресурсов

А.В. Олейникова, А.А. Олейников
Астраханский государственный технический университет,
414056, г. Астрахань, ул. Татищева, 16, Россия

Резюме. Цель. Добыча водных биологических ресурсов является важным сектором экономики РФ. Наряду с достижением высоких экономических показателей, на которые нацелены заготовители, на первом плане кроме добычи выделяется и требование к сохранению популяций и поддержанию разнообразия видов водных биологических ресурсов. Достижение целей, устанавливаемых на основе перечисленных требований возможно за счет повышения эффективности обработки информации, получаемой из оперативных отчетов собираемых по отделам территориальных управлений. Отсутствие автоматизации обработки электронных таблиц и формирование аналитических отчетов в ручном режиме затрудняет формирование четкой обобщенной картины по состоянию отрасли. **Метод**. В настоящем исследовании использовались методы интеллектуального анализа данных анализа, анализа статистических данных. **Результат**. Разработана модель интеллектуального анализа данных для исследования системы квот вылова водных биологических ресурсов созданная в программе Loginom. **Вывод**. Разработанная модель позволяет не только оценить уровни освоения квот, но и производить прогнозирование уровеней освоения с применением модуля arimax, а также показывает возможность создания полноценной информационной системы учета добычи водных биологических ресурсов на ее основе.

Ключевые слова: водные биологические ресурсы, квота, уровень освоения, модель интеллектуального анализа данных

Для цитирования: А.В. Олейникова, А.А. Олейников. Модель интеллектуального анализа данных для исследования системы квот вылова водных биологических ресурсов. Вестник Дагестанского государственного технического университета. Технические науки. 2023; 50(4): 121-128. DOI:10.21822/2073-6185-2023-50-4-121-128

Data mining model for the study of the system of quotas for the catch of aquatic biological resources

A.V. Oleynikova, A.A. Oleynikov
Astrakhan State Technical University,
16 Tatishcheva St., Astrakhan 414056, Russia

Abstract. Objective. Extraction of aquatic biological resources is an important sector of the Russian economy. Along with the achievement of high economic indicators, which are aimed at harvesters, in the foreground, in addition to production, the requirement to preserve populations and maintain the diversity of aquatic biological resources is also highlighted. Achieving the goals set on the basis of the listed requirements is possible by increasing the efficiency of processing information obtained from operational reports collected by departments of territorial administrations. The lack of automation in the processing of spreadsheets and the generation of analytical reports in manual mode makes it difficult to form a clear generalized picture of the state of the industry. **Method**. In the present study, methods of data mining analysis, analysis of statistical data were used. **Result**. A data mining model has been developed to study the system of quotas for catching aquatic biological

resources, created in the Loginom program. **Conclusion**. The developed model allows not only to estimate the levels of development of quotas, but also to predict the levels of development using the arimax module, and also shows the possibility of creating a full-fledged information system for accounting for the extraction of aquatic biological resources based on it.

Keywords: aquatic biological resources, quota, development level, data mining model.

For citation: A.V. Oleynikova, A.A. Oleynikov. Data mining model for the study of the system of quotas for the catch of aquatic biological resources. Herald of Daghestan State Technical University. Technical Sciences. 2023; 50(4): 121-128. DOI:10.21822/2073-6185-2023-50-4-121-128

Введение. Волго-Каспийское территориальное управление Федерального агентства по рыболовству (Управление) является территориальным органом Федерального агентства по рыболовству, которое создано для осуществления функций по контролю (надзору) в области рыболовства и сохранения водных биологических ресурсов на водных объектах рыбохозяйственного значения Республики Калмыкия, Астраханской области, Волгоградской области, Саратовской области, государственному надзору за торговым мореплаванием в части обеспечения безопасности плавания судов рыбопромыслового флота в районах промысла при осуществлении рыболовства, по оказанию государственных услуг, управлению государственным имуществом в сфере рыбохозяйственной деятельности, охраны, рационального использования, изучения, сохранения, воспроизводства водных биологических ресурсов и среды их обитания, а также рыбоводства (аквакультуры), товарного рыбоводства, производства рыбной и иной продукции из водных биологических ресурсов, обеспечения безопасности плавания судов рыбопромыслового флота и аварийно-спасательных работ в районах промысла при осуществлении рыболовства, производственной деятельности на судах рыбопромыслового флота и в морских портах в отношении морских терминалов, предназначенных для комплексного обслуживания судов рыбопромыслового флота.

В сферу деятельности Управления входит также и управление водными биоресурсами, которое осуществляется с помощью контроля за распределением квот добычи, отслеживание нераспределенных остатков квот каждого пользователя между его структурными подразделениями с целью недопущения превышения квот, выделенных приказом, по каждому району, виду водных биоресурсов и типу квоты.

Постановка задачи. В условиях рыночной экономики необходимо минимизировать количество и упростить структуру информации таким образом, чтобы сбор данных не мешал ведению предпринимательской деятельности. Необходима разработка алгоритма системного анализа рыбохозяйственной информации, формирования стандартных аналитических форм, обеспечивающих не только суммирование уловов, но и поддержку принятия решений в области контроля за состоянием водных биоресурсов, рыбопромысловой деятельности и управления рыболовством. В настоящее время распространенным механизмом повышения эффективности работы Управления, ускорения решения проблем и повышения контроля ее деятельности, благодаря которому открываются новые подходы к решению стандартных задач, является система поддержки принятия решений (СППР). В качестве основной идеи реализации СППР предлагается использование аналитической платформы Loginom, которая является основой для создания библиотеки компонентов в области анализа данных [1,4].

Цель работы заключается в разработке модели интеллектуального анализа данных для исследования системы квот вылова водных биологических ресурсов.

Методы исследования. Для анализа предметной области, разработки библиотеки компонентов и данных по системе квот вылова (фактического освоения) водных биологических ресурсов (ВБР) Волго-Каспийского территориального управления Федерального агентства по рыболовству предлагается использовать аналитическую платформу Loginom. Loginom предоставляет совершенно новый подход к проектированию, который ориентирован на потребности хозяйствующих субъектов разных уровней и позволяет строить модели для

интеллектуального анализа данных разной степени сложности. Применение Loginom для интеллектуального анализа данных позволяет осуществлять проектирование сверху вниз, начиная постановкой задачи исследования и заканчивая решением отдельных подзадач.

В самом начале после описания решения поставленной задачи происходит выделение независимых друг от друга подзадач. Далее аналогичным образом для формирования следующего уровня описывают подзадачи, уточняя на каждом шаге функционал проектируемого решения. Данный процесс происходит до момента выделения подзадач, решение которых является наиболее простым и очевидным и не требует дальнейшей детализации. Loginom предоставляет возможность создания сценария анализа данных, включающего все необходимые методы. В сценарий входят комбинации различных узлом обработки данных, реализующих те или иные методы. Их настраивает пользователь в соответствии с задачами исследования. А затем каждый узел в составе сценария выполняет свою функцию. Проблемам интеллектуального анализа данных посвящены работы Васильева В.В. [2], Дорофеюк А.А. [5], Зеленкова Ю.А. [6], Семенова В.А. [8] и др.

Обсуждение результатов. Рассмотрим более подробно работу компонентов аналитического модуля на примере системы квотирования водных биологических ресурсов отдела контроля, надзора, охраны ВБР и среды обитания по Астраханской области, отдела контроля, надзора, охраны ВБР и среды обитания по Саратовской области, отдела контроля, надзора, охраны ВБР и среды обитания по Республике Калмыкия и Нижневолжского отдела контроля, надзора, охраны ВБР и среды обитания (г. Волгоград) за 2019 г. (рис.1.).

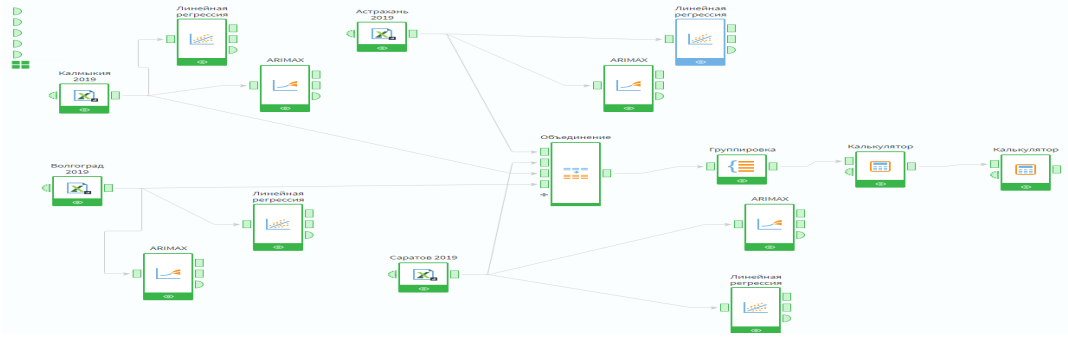

Рис. 1. Архитектура компонентов
Fig. 1. Component Architecture

Для установления зависимости между фактическим освоением в тоннах и квотой на разращённый объем добычи (вылова) ВБР за каждый месяц 2019 г. по отделам Управления и по видам ВБР использовали компонент «Линейная регрессия». Регрессионный анализ помогает понять, как «типичное» значение зависимой переменной изменяется при изменении одной из независимых переменных, в то время как другие независимые переменные остаются фиксированными [3,7-10].

Регрессионная модель:

$$y = f(x, b) + \varepsilon, E\ (\varepsilon),$$

где, b-параметры модели, ε −случайная ошибка модели; называется линейной регрессией, если функция регрессии $f(x, b)$ имеет вид

$$f(x, b) = b_0 + b_1 x_1 + b_2 x_2 + \cdots\ b_k x_k,$$

где b_j-параметры (коэффициенты) регрессии, x_j-регрессоры (факторы модели), *k* — количество факторов модели.

Коэффициенты линейной регрессии показывают скорость изменения зависимой переменной по данному фактору, при фиксированных остальных факторах (в линейной модели эта скорость постоянна):

$$\forall_j \quad b_j = \frac{\partial f}{\partial x_j} = const$$

Параметр b_0 , при котором нет факторов, называют часто *константой*.

Формально — это значение функции при нулевом значении всех факторов. Для аналитических целей удобно считать, что константа — это параметр при «факторе», равном 1 (или другой произвольной постоянной, поэтому константой называют также и этот «фактор»). В таком случае, если перенумеровать факторы и параметры исходной модели с учетом этого (оставив обозначение общего количества факторов — k), то линейную функцию регрессии можно записать в следующем виде, формально не содержащем константу:

$$f(x,b) = b_1x_1 + b_2x_2 + \cdots b_kx_k = \sum_{j=1}^{k} b_jx_j = x^Tb,$$

где $x^T = (x_1 + x_2, \ldots\ldots, x_k)$-вектор регрессоров, $b = (b_1 + b_2, \ldots\ldots, b_k)^T$-вектор-столбец параметров (коэффициентов).

При использовании линейной регрессии, была установлена зависимость, при которой объем фактического освоения квот от промышленного рыболовства в большей степени влияет на объем выделения квот по отделу контроля, надзора, охраны ВБР и среды обитания Астраханской области. Визуализатор Отчет по регрессии отображает параметры и результаты статистических тестов для анализа регрессионной модели (рис.2).

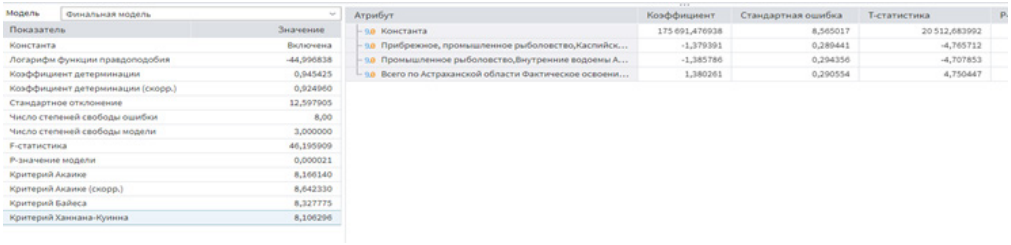

Рис. 2. Отчет по регрессии Астраханской области
Fig. 2. Report on regression of the Astrakhan region

Общее качество регрессионной модели определяется с помощью коэффициента детерминации R^2. В отчете по регрессии Информация о модели его величина составляет 0,945. Поскольку максимальное значение R^2=1, что можно утверждать, что качество регрессионной модели весьма высокое (предлагаемая модель объясняет около 94,5 % дисперсии результативной переменной).

Согласно графикам линейной регрессии по уровню освоение квот можно сделать вывод, что на объем выделения квот по Астраханской области за 2019 г. влияет в большей степени объем фактического освоения квот по промышленному виду рыболовства (рис.3-4.).

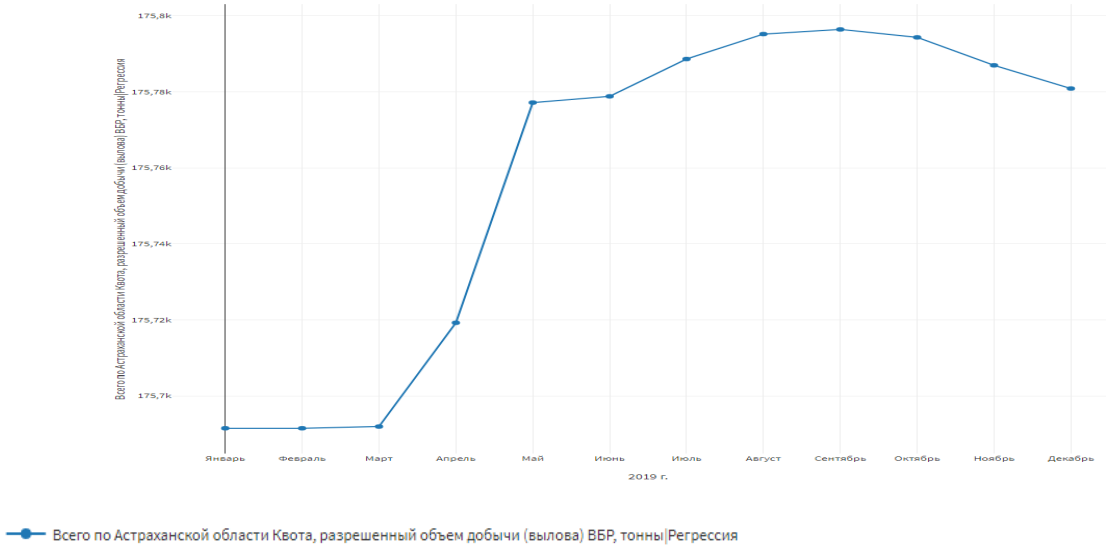

Рис. 3. График регрессии по Астраханской области
Fig. 3. Regression graph for the Astrakhan region

Рис. 4. Фактическое освоение квот по видам рыболовства по Астрахани
Fig. 4. Actual utilization of quotas by type of fishing in Astrakhan

Параллельно был использован обработчик ARIMAX, с помощью которого были получены прогнозные значения объемов выделенных квот по Астраханской области. ARIMAX математическая модель для анализа временных рядов, объединяющая в себе интегрированную авторегрессию, скользящее среднее и возможность учета дополнительных внешних факторов [11]. Результирующее множество сравнено с данными фактического потребления и увидеть, при этом результат прогнозирования на основе алгоритма ARIMAX оказался более приближен к реальному (рис.5.).

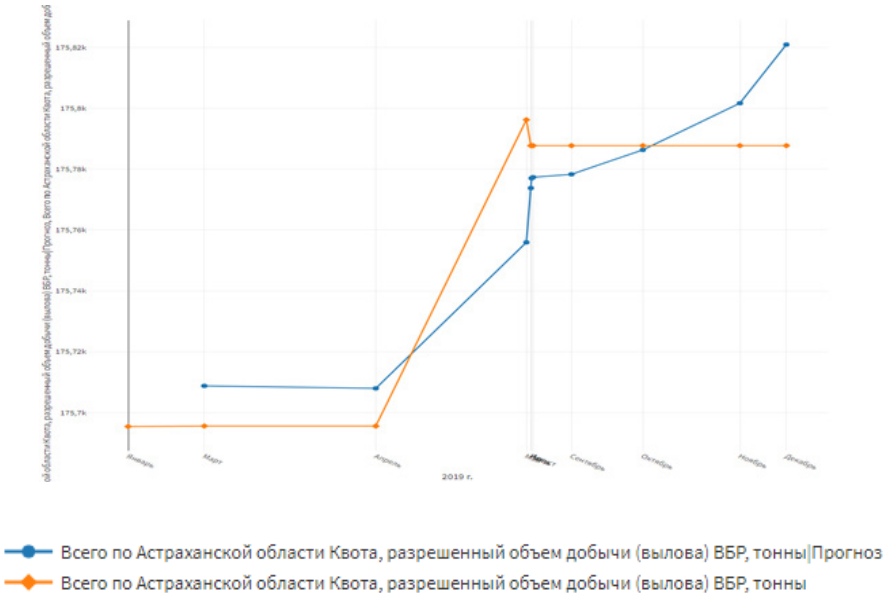

Рис. 5. Прогноз выделения квот на основе алгоритма ARIMAX по Астраханской области
Fig. 5. Forecast of quota allocation based on the ARIMAX algorithm for the Astrakhan region

Аналогичные выводы были сделаны при анализе выделения квот и их фактическим освоением за 2019 год по видам рыболовства и по отделам Волго-Каспийского территориального управления Федерального агентства по рыболовству, находящимся в

Саратовской области, Волгоградской области и Республике Калмыкия в виду схожести данных.

Так как значительное влияние на объем выделения квот по отделам Управления оказывает именно фактическое освоения промышленного рыболовства были взяты показатели фактического освоения квот по промышленному рыболовству для дальнейшего проведения анализа уровня поступления квот именно по отделам Управления. Для того чтобы выявить значимость (влияние) отдела необходимо узнать общий объем выделения квот, общий объем фактического поступления квот, общий объем освоения квот в целом по Управлению, а для этого используем компонент Loginom Объединение. После объединения значений по всем отделам Управления, производим настройку полей компонента Группировка, после чего получаем весь объем набора данных по Управлению, по отделам, по промышленному виду рыболовства, по каждому месяцу 2019 года. С помощью компонента трансформации Калькулятор, производим расчет общего объёмы выделенных квот по Управлению, общего объема фактического освоения квот по Управлению, общего процента освоения по Управлению. В дальнейшем еще раз используем компонент трансформации Калькулятор для выявления значимого отдела по уровню освоения квот, проценту освоения квот и подводим итоги.

Согласно полученным данным и представленному визуализатору диаграмма столбчатая приходим к выводу, что целесообразно больше выделять квот именно на промышленное рыболовство по Астраханской области (рис.6), так как в дальнейшем от желаемого уровня выделения квот зависит обеспеченность продовольственной безопасности региона, стабильность социально-экономических процессов.

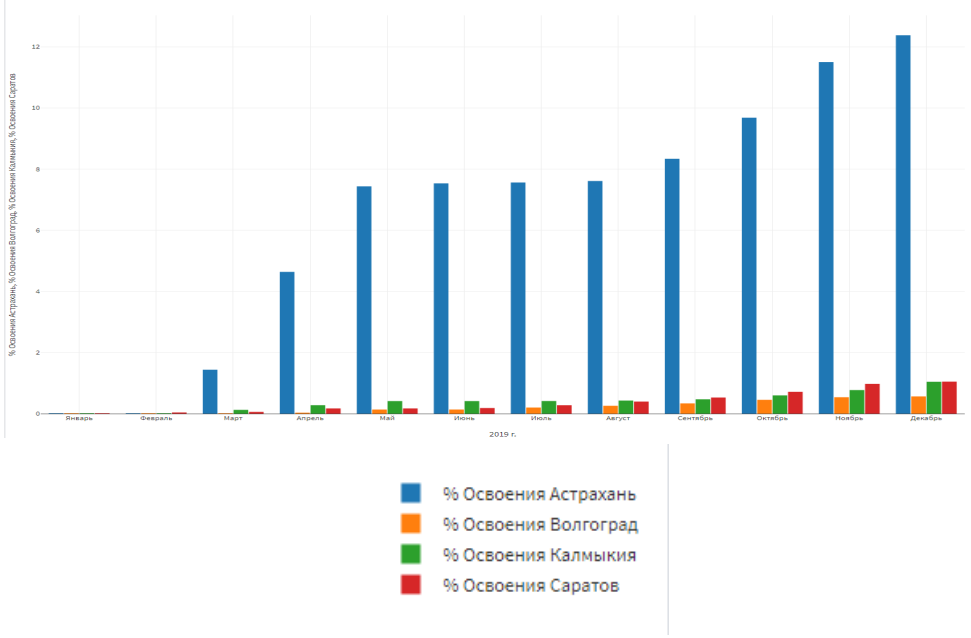

Рис. 6. Диаграмма столбчатая по % освоения квот Управления
Fig. 6. Column chart for % of Management quota development

Традиционно добыча и переработка рыбы являлись важным средством обеспечения занятости населения Астраханской области. Поэтому резкое сокращение объемов промышленного вылова рыбы (и, как следствие, объемов ее промышленной переработки, объемов деятельности рыботорговых компаний и пр.) значительно повлияло на занятость населения (особенно в сельских районах области). Городским населением АО вылов рыбы осуществляется, в основном, в качестве одного из видов рекреационного поведения, а не как средство получения дохода или пополнения рационов. Причина – рыбу-сырец обычно дешевле купить, чем выловить самостоятельно (с учетом субъективных оценок денежных потерь, соответствующих расходам времени на обеспечение вылова и обработки рыбы, а также факторов неопределенности в отношении предполагаемых результатов любительского лова и пр.). В тоже время для сельского населения любительский лов рыбы во внутренних

водоемах и (частично) ее переработка, являются важным средством обеспечения личных пищевых рационов; получения доходов от продажи рыбы-сырца или переработанной. Рыбалка на территории Астраханской (наряду с охотой) традиционно является важным средством привлечения въездного турконтингента на территорию области.

Вывод. В ходе исследования была проанализирована предметная область с применением метода интеллектуального анализа данных. Для проведения интеллектуального анализа данных применялась программа Loginom.

По мере возрастания объёма данных упрощается рассмотрение их как стохастического процесса с некоторой тенденцией. Выявление закономерностей в этих данных может иметь большой потенциал благодаря комплексному применению методов анализа данных. Методы интеллектуального анализа данных основываются на таких концептуальных направлениях как математическая статистика, машинное обучение и искусственный интеллект. При этом точность выявленных моделей будет возрастать с ростом объема и разнотипности данных.

Благодаря применению данных систем можно производить визуализацию и прогнозирование в регионах не только объема выделения квот, уровня и процента освоения, но и прогнозировать социально-экономическую ситуацию в регионах. При этом с помощью интеллектуального анализа данных можно извлечь полезную информацию из больших объемов данных, которые могут быть использованы для принятия экономически оптимальных решений, повышения эффективности и качества работы в организационной структуре.

Библиографический список:

1. Абакумов Е.М. Сравнительный анализ алгоритмов интеллектуального анализа данных для обработки неструктурированных данных / Е. М. Абакумов, А. В. Горлов, М. Д. Кириллова, О. В. Ульянин // Физико-технические интеллектуальные системы (ФТИС-2022): Сборник трудов научно-практической конференции, Москва, 08–10 февраля 2022 года. – Москва: Буки Веди, 2022. – С. 101-106.
2. Васильев, В. В. Анализ существующих методов интеллектуального анализа неформализованных данных / В. В. Васильев // Международная конференция по мягким вычислениям и измерениям. – 2009. – Т. 1. – С. 99-102.
3. Велигура, А. В. Анализ существующих методологий интеллектуального анализа данных для производственных систем / А. В. Велигура, Э. К. Мусаева // Вестник Луганского национального университета имени Владимира Даля. – 2020. – № 5(35). – С. 45-50.
4. Джамансариев, Н. Б. Сравнительный анализ программ для интеллектуального анализа данных / Н. Б. Джамансариев // Информационные технологии в науке, управлении, социальной сфере и медицине: Сборник научных трудов II Международной конференции, Томск, 19–22 мая 2015 года / Национальный исследовательский Томский политехнический университет. – Томск: Национальный исследовательский Томский политехнический университет, 2015. – С. 135-136.
5. Дорофеюк, А. А. Методы и алгоритмы интеллектуального анализа сложно организованных данных / А. А. Дорофеюк, Ю. А. Дорофеюк // Управление развитием крупномасштабных систем: Современные проблемы / Под редакцией А.Д. Цвиркуна. Том Выпуск 2. – Москва: Общество с ограниченной ответственностью Издательская фирма «Физико-математическая литература», 2015. – С. 378-409.
6. Зеленков, Ю. А. Динамика исследований в области интеллектуального анализа данных: тематический анализ публикаций за 20 лет / Ю. А. Зеленков, Е. А. Анисичкина // Бизнес-информатика. – 2021. – Т. 15, № 1. – С. 30-46.
7. Рогоза, М. В. Многомерный анализ агрегированной численной информации с применением интеллектуального анализа данных / М. В. Рогоза // Актуальные проблемы теории и практики развития экономики региона: Материалы 2-й межвузовской студенческой научно-практической конференции, Калуга, 21 апреля 2015 года. – Калуга: Алькор Паблишерс, 2015. – С. 448-451.
8. Семенов, В. А. Анализ методов и задач интеллектуального анализа данных / В. А. Семенов, М. М. Семенова, Л. В. Мызникова // Информационные технологии, системный анализ и управление (Итсау-2021): сборник трудов XIX Всероссийской научной конференции молодых ученых, аспирантов и студентов, Таганрог, 09–11 декабря 2021 года. – Ростов-на-Дону - Таганрог: Южный федеральный университет, 2021. – С. 209-212. – EDN GNFPQR.
9. Татур, М. М. Анализ временных рядов как элемент процесса интеллектуального анализа данных обстановки с техногенными пожарами / М. М. Татур, А. Г. Иваницкий, В. М. Проровский // Чрезвычайные ситуации: предупреждение и ликвидация. – 2021. – № 1(49). – С. 56-68.

10. Mikheyenkova, M. A. On the Approach to Intelligent Data Analysis in the Social Sciences / M. A. Mikheyenkova // Открытые семантические технологии проектирования интеллектуальных систем. – 2020. – No. 4. – P. 129-132.
11. А.В. Олейникова, А.А. Олейников. Онтологическая модель знаний как основа автоматизации процесса распределения квот добычи водных биологических ресурсов и контроля за уровнем их освоения. Вестник Дагестанского государственного технического университета. Технические науки. 2023; 50(2):126-133. DOI:10.21822/2073-6185-2023-50-2-126-133

References:

1. Abakumov E.M, Gorlov A. V., Kirillova M. D., Ulyanin O. V. Comparative analysis of data mining algorithms for processing unstructured data. *Physical and technical intelligent systems (FTIS-2022): Collection of scientific papers -practical conference*, Moscow, February 08–10, 2022. - Moscow: Buki Vedi, 2022;101-106. (In Russ)
2. Vasiliev, V. V. Analysis of existing methods of data mining. *International Conference on Soft Computing and Measurements*. 2009; 1: 99-102. (In Russ)
3. Veligura, A. V., Musaeva E. K. Analysis of existing data mining methodologies for production systems. *Bulletin of the Volodymyr Dahl Lugansk National University*. 2020;5 (35): 45-50. (In Russ)
4. Dzhamansariev, N. B. Comparative analysis of programs for data mining. *Information technologies in science, management, social sphere and medicine: Collection of scientific papers of the II International Conference*, Tomsk, May 19–22, 2015.National Research Tomsk Polytechnic University. - Tomsk: National Research Tomsk Polytechnic University, 2015; 135-136. (In Russ)
5. Dorofeyuk, A.A., Dorofeyuk Yu.A. Methods and algorithms for data mining of complexly organized data. Development management of large-scale systems: Modern problems. *Edited by A.D. Tsvirkun. Volume Issue 2. - Moscow: Limited Liability Company Publishing Company "Physical and Mathematical Literature*", 2015; 378-409. (In Russ)
6. Zelenkov, Yu. A.,Anisichkina E. A. Dynamics of research in the field of data mining: thematic analysis of publications for 20 years. *Business Informatics*. 2021; 15(1): 30-46. (In Russ)
7. Rogoza, M. V. Multivariate analysis of aggregated numerical information using data mining. *Actual problems of the theory and practice of regional economic development*: Proceedings of the 2nd interuniversity student scientific and practical conference, Kaluga, April 21 2015. - Kaluga: Alcor Publishers, 2015; 448-451. (In Russ)
8. Semenov, V. A., Semenova M. M., Myznikova L. V. Analysis of methods and tasks of data mining. Information technologies, system analysis and management (Itsau-2021): collection of works XIX All-Russian scientific conference of young scientists, graduate students and students, Taganrog, December 09–11, 2021. - Rostov-on-Don - Taganrog: Southern Federal University, 2021; 209-212. (In Russ)
9. Tatur, M. M., Ivanitsky, A. G., Prorovsky, V. M. Time series analysis as an element of the process of data mining of the situation with man-made fires. *Emergencies: prevention and liquidation*. 2021;1 (49):56-68. (In Russ)
10. Mikheyenkova, M. A. On the Approach to Intelligent Data Analysis in the Social Sciences / M. A. Mikheyenkova. Открытые семантические технологии проектирования интеллектуальных систем. 2020; 4:129-132. (In Russ)
11. A.V. Oleynikova, A.A. Oleynikov. Ontological model of knowledge as the basis for automation of the process of distribution of quotas for the extraction of aquatic biological resources and control over the level of their development. Herald of Daghestan State Technical University. Technical Science. 2023; 50(2):126-133. DOI:10.21822/2073-6185-2023-50-2-126-133(In Russ)

Сведения об авторах:

Олейникова Алла Владимировна, магистрант кафедры прикладная информатика; a.oleynikova.astu@mail.ru

Олейников Александр Александрович, кандидат технических наук, доцент, кафедра систем автоматизированного проектирования и моделирования; aleksandr6333@gmail.com

Information about authors:

Alla V. Oleinikova, Undergraduate of the Department of Applied Informatics, a.oleynikova.astu@mail.ru

Alexander A. Oleinikov, Cand.Sci. (Eng.), Assoc. Prof., Department of Computer-Aided Design and Modeling Systems; aleksandr6333@gmail.com

Конфликт интересов/Conflict of interest.

Авторы заявляют об отсутствии конфликта интересов/The authors declare no conflict of interest.

Поступила в редакцию/Received 10.09.2023.

Одобрена после рецензирования/ Reviced 01.10.2023.

Принята в печать/Accepted for publication 01.10.2023.