УДК 004.415.28 **DOI: 10.31799/2077-5687-2022-1-20-37**

СПЕЦИАЛИЗИРОВАННЫЕ ИНСТРУМЕНТЫ ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ DATA MINING

Н. В. Богословская, А. В. Бржезовский, О. И. Красильникова
Санкт-Петербургский государственный университет аэрокосмического приборостроения

Обоснована актуальность использования технологии визуализации при интерпретации результатов анализа данных. Рассмотрены практические аспекты визуализации результатов Data Mining. Приведены классификации аналитических методов и существующих практических задач. Предложена классификация инструментов визуализации в разрезе аналитических методов и задач аналитики данных. Сделан обзор современных аналитических систем, позволяющих получить визуальное представление о данных. Обоснован выбор аналитических систем, визуальные компоненты которых используются в исследовании. Приведено краткое описание задач аналитики. Выполнен анализ встроенных визуализаторов для различных аналитических систем с учетом методов и задач аналитики. Рассмотрены совпадения и отличия в используемых встроенных инструментах визуализации для каждой из систем. Обозначена проблема отсутствия визуальных компонентов для интерпретации результатов модели «ассоциативные правила». Отмечены современные средства визуализации на примере интерактивных панелей «Профили кластеров» в пакете Loginom. Выполнен анализ специализированных средств для визуализации метрик в рассмотренных моделях Data Mining. Сформулированы выводы о возможности дополнения встроенных средств конкретной аналитической системы альтернативными компонентами других систем.

Ключевые слова: визуальное представление данных, аналитические платформы, Data Mining, инструменты интерпретации результатов моделирования, классификация задач Data Mining, дескриптивная аналитика, предикативная аналитика, ассоциативные правила, кластеризация, классификация, регрессионный анализ, IBM SPSS, Loginom, 1С: Предприятие.

***Для цитирования**:*
Богословская Н. В., Бржезовский А. В., Красильникова О. И. Специализированные инструменты визуализации результатов data mining // Системный анализ и логистика: журнал.: выпуск №1(31), ISSN 2007-5687. – СПб.: ГУАП., 2022 – с.20-37. РИНЦ. DOI: 10.31799/2077-5687-2022-1-20-37.

SPECIALIZED TOOLS FOR VISUALIZING DATA MINING RESULTS

N. V. Bogoslovskaia, A. V. Brzhezovskii, O. I. Krasilnikova
St. Petersburg State University of Aerospace Instrumentation

The relevance of the use of visualization technology in interpreting the results of data analysis is substantiated. Practical aspects of visualization of Data Mining results are considered. Classifications of analytical methods and existing practical tasks are given. The classification of visualization tools in the context of analytical methods and data analytics tasks is proposed. An overview of modern analytical systems allowing to get a visual representation of the data is made. The choice of analytical systems, the visual components of which are used in the study, is justified. A brief description of the tasks of analytics is given. The analysis of the built-in visualizers for various analytical systems is carried out taking into account the methods and tasks of analytics. The similarities and differences in the built-in visualization tools used for each of the systems are considered. The problem of the lack of visual components for interpreting the results of the "associative rules" model is indicated. Modern visualization tools are noted on the example of interactive panels "Cluster Profiles" in the Loginom package. The analysis of specialized tools for visualization of metrics in the considered Data Mining models is carried out. Conclusions are formulated about the possibility of supplementing the built-in tools of a specific analytical system with alternative components of other systems.

Keywords: visual representation of data, analytical platforms, Data Mining, tools for interpreting modeling results, classification of Data Mining tasks, descriptive analytics, predicative analytics, associative rules, clustering, classification, regression analysis, IBM SPSS, Loginom, 1C: Company.

For citation:
Bogoslovskaia N. V., Brzhezovskii A. V., Krasilnikova O. I. Specialized tools for visualizing data mining results // System analysis and logistics.: №1(31), ISSN 2007–5687. – Russia, Saint-Petersburg.: SUAI., 2022 –p 20-37. DOI: 10.31799/2077-5687-2022-1-20-37.

Введение

Известный американский специалист по математической статистике Джон Тьюки был убежден, что многое можно узнать из данных, просто визуализируя их. Важнейшим элементом аналитики Дж. Тьюки определил широкое использование визуального представления многомерных данных [1]. За последние годы в этой области было выполнено немало исследований, направленных на разработку новых методов визуализации и решение целого ряда связанных с этим проблем [2].

В современных аналитических системах и Data Sience активно используют десятки различных методов визуализации. Традиционно средства визуализации классифицируют как табличные и графические, одномерные и многомерные, общего и специализированного назначения. Такая классификация отражает характер данных и специфику решаемых задач, связанных с обработкой данных, например, в таких программных средствах, как офисные приложения [3].

В аналитических платформах инструменты визуализации являются обязательной частью сценария обработки, а не только инструментом интерпретации результата. В системах аналитики визуализаторы не относятся к отдельным подсистемам, они встроены в каждый компонент системы. Классификацию визуализаторов принято рассматривать в контексте жизненного цикла обработки данных:

- визуализация источника данных,
- визуализация загруженной выборки,
- визуализация результатов предобработки данных,
- визуализация промежуточных результатов,
- визуализация результатов анализа.

Авторы статьи, учитывая, что использование аналитических инструментов обеспечивает аналитика средствами проверки гипотез, предлагают альтернативную классификацию визуализаторов в разрезе используемых аналитических методов и категорий решаемых задач.

При выдвижении гипотез аналитик опирается на собственный опыт, но проверка гипотезы должна быть основана на знаниях, накопленных в исследуемых данных. Эти знания не лежат на поверхности, они скрыты от пользователя, простые инструменты визуализации не позволяют обнаружить в «сырых» данных ранее неизвестные, нетривиальные, практически полезные и доступные интерпретации знания, необходимые для принятия решений в различных сферах человеческой деятельности. Специальные методы автоматического анализа, при помощи которых добываются знания из огромных массивов информации, принято называть Data Mining [4].

Классификация задач Data Mining позволяет сгруппировать аналитические методы в дескриптивные и предикативные. Дескриптивная аналитика реализуется в задачах ассоциативных правил и задачах кластеризации; предикативная аналитика реализуется в задачах классификации и регрессионного анализа. Соотношение методов, задач и особенностей извлекаемых знаний приведено в табл.1.

Таблица 1 – Классификация задача Data Mining

Аналитические методы	Задачи аналитики	Знания
Дескриптивная аналитика	Ассоциативные правила Кластеризация	Какие взаимосвязи между характеристиками? Какие события происходят одновременно? Какова структура клиентской базы? Какой профиль идеального клиента?
Предикативная аналитика	Классификация	Откликнется ли клиент на данную маркетинговую кампанию?

	Регрессия	Какие из потенциальных клиентов, вероятно, совершат приобретение услуги в следующем месяце? Какой размер прибыли будет в следующем месяце? Какой прогнозируемый спрос на товар на следующий месяц?

Необходимость обработки как структурированных, так и слабоструктурированных, неструктурированных массивов данных привела к созданию и применению специальных высокопроизводительных инструментов и методов анализа данных. В табл.2 показано развитие классификации задач Data Mining.

Дескриптивная аналитика является стадией подготовки данных для дальнейшего анализа. Выделение методов анализа в диагностическую аналитику связано с практическим ракурсом оценки современных задач аналитики. Диагностическую аналитику можно рассматривать как форму расширенной аналитики для базовой описательной. Интерпретация найденных знаний имеет цель получения важных для бизнес-деятельности сведений, которые не только фиксируют сложившуюся ситуацию, но и позволяют ответить на вопрос «Почему это случилось?».

Уточняя расширенную классификацию задач аналитики, можно позиционировать задачи кластерного анализа как задачи, соотнесенные с диагностической аналитикой. Кластерный анализ – группа методов, используемых для классификации объектов или событий в относительно гомогенные группы, позволяющие получить дополнительную информацию о том, через какие объекты классы взаимосвязаны друг с другом [5]. С точки зрения диагностической аналитики, кластеризация помогает выявлять (группировать) аномалии и случайные связи между событиями и действиями.

Таблица 2 – Современные задачи аналитики

Методы анализа	Задачи аналитики
Дескриптивная аналитика *Что случилось?*	Ассоциативные правила
Диагностическая аналитика *Почему это случилось?*	Кластерный анализ
Предикативная аналитика *Что может случиться?*	Классификация Регрессионный анализ
Предписывающая аналитика *Что делать?*	Машинное обучение Симуляция Нейронные сети

Материалы и методы исследования

Для процесса анализа данных существует множество инструментов бизнес-аналитики (BI-инструментов). Современные аналитические системы представляют собой сложные, комплексные проекты, реализуемые в процессе цифровизации бизнеса, позволяющие анализировать данные и визуализировать результаты.

Ряд BI-инструментов дают возможность пользователю быстро получить представление о данных. Существует несколько API, которые позволяют относительно легко и с большей

эффективностью выполнять конкретную аналитику. Например: KNIME, Rapid Miner, Splunk, TIBCO Spotfire, Qlik, SQL, MS Excel [6].

Предикативная и предписывающая аналитики используют модели прогнозного моделирования, а определенные статистические и программные инструменты позволяют пользователю создавать такие модели с использованием библиотек, которые помогают в разработке сложных статистических моделей, моделей машинного обучения и глубокого обучения. Популярными инструментами реализации этих моделей являются: Python, R, SAS [7].

Одним из наиболее важных аспектов аналитики является представление сложной информации в удобном для понимания визуальном формате. Различные инструменты помогают в создании графических отчетов, упрощая процесс их создания. Популярными средствами аналитической отчетности и визуализации являются: Tableau, MS Excel, Power BI, Chartio, Redash, Google Data Studio, Loginom [6].

Авторы статьи, являясь преподавателя вуза, анализируют программные средства визуализации результатов аналитической обработки данных для применения в учебных дисциплинах. В статье [8] были рассмотрены визуализаторы общего назначения. Следует отметить, что при изучении сложных методов анализа интерпретация полученных результатов общими средствами визуализации становится неэффективной.

В представленной статье рассматриваются специализированные средства компьютерной графики в аналитических системах с учетом возможностей и потребностей учебного процесса вуза. Выбор аналитических платформ обоснован, с одной стороны, их использованием в различных учебных дисциплинах образовательной программы 09.03.02(06) Информационные системы и технологии в бизнесе, с другой – принадлежностью к различным категориям ПО. Loginom (Россия) - Low-code платформа для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации [9]. IBM SPSS (США) - компьютерная программа для статистической и другой обработки данных, один из лидеров рынка в области коммерческих статистических продуктов [10]. 1С:Предприятие (Россия) – комплекс интегрируемых решений автоматизации предприятий, располагающий встроенным механизмом анализа данных и прогнозирования [11].

Продолжая классификацию задач аналитики следует рассмотреть используемые специализированные инструменты визуализации при интерпретации результатов анализа. В табл. 3 рассмотрены примеры встроенных визуализаторов в различных аналитических системах: аналитических платформах Loginom и IBM SPSS Modeler, системе со встроенной аналитикой данных 1С: Предприятие.

Таблица 3 –Задачи аналитики и доступные встроенные типы визуализации

Методы анализа	Задачи аналитики	Встроенные визуализаторы		
		Loginom Loginom company, Россия	1С: Предприятие фирма «1С», Россия	IBM SPSS Modeler IBM, США
Дескриптивная аналитика	Ассоциативные правила	Таблица правил	Таблица правил	Таблица правил связывания; График Web; График Directed Web
Диагностическая аналитика	Кластерный анализ	Интерактивная таблица; Профили кластеров	Таблица кластеров; Дендрограмма	Интерактивная таблица; Диаграмма кластеров

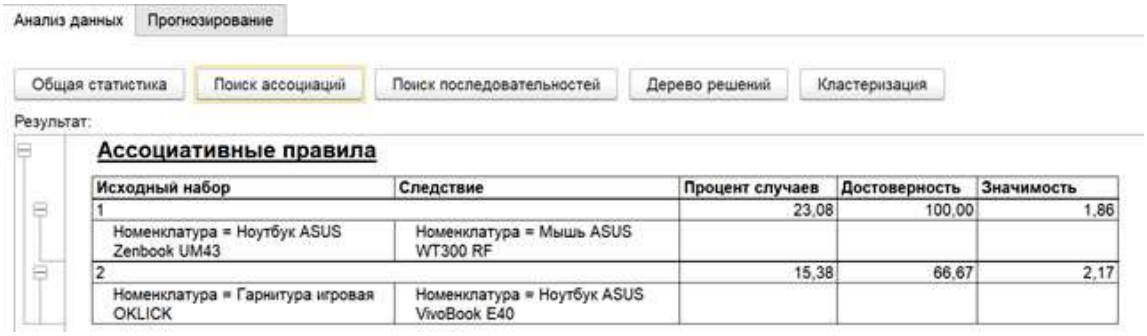
Предикативная аналитика	Классификация Регрессионный анализ	Таблица; Дерево регрессии	Дерево решений	Таблица; Деревья классификации Деревья регрессии

Результаты исследования и их обсуждение

Методы анализа, представленные в табл. 3, отличаются уровнем сложности работы с информацией и степенью человеческого участия. Из опыта работы с различными аналитическими системами авторы выдвигают предположение, что усложнение моделей от описательной к предсказательной аналитике сопровождается увеличением разнообразия и сложности средств визуализации как при интерпретации результатов, так и при оценке качества исследования, при этом «простые» модели аналитики имеют недостаточную поддержку средствами визуализации и оценки качества.

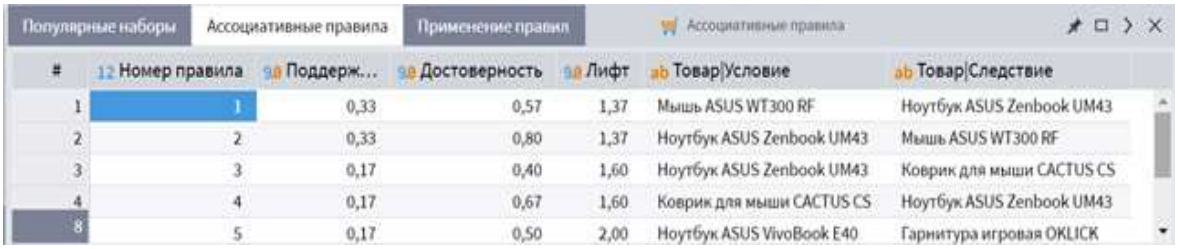
Дескриптивная аналитика. Ассоциативные правила

В изучаемых системах: Loginom, 1С: Предприятие, IBM SPSS Modeler визуализация результатов модели «Ассоциативные правила» выполнена инструментом Таблица. Структура таблиц выглядит примерно одинаково: *Условие* правила, *Следствие* правила, метрики оценки (рис.1, 2, 3).



Исходный набор	Следствие	Процент случаев	Достоверность	Значимость
1		23,08	100,00	1,86
Номенклатура = Ноутбук ASUS Zenbook UM43	Номенклатура = Мышь ASUS WT300 RF			
2		15,38	66,67	2,17
Номенклатура = Гарнитура игровая OKLICK	Номенклатура = Ноутбук ASUS VivoBook E40			

Рис.1. Таблица ассоциативных правил в системе 1С



#	Номер правила	Поддерж...	Достоверность	Лифт	Товар|Условие	Товар|Следствие
1	1	0,33	0,57	1,37	Мышь ASUS WT300 RF	Ноутбук ASUS Zenbook UM43
2	2	0,33	0,80	1,37	Ноутбук ASUS Zenbook UM43	Мышь ASUS WT300 RF
3	3	0,17	0,40	1,60	Ноутбук ASUS Zenbook UM43	Коврик для мыши CACTUS CS
4	4	0,17	0,67	1,60	Коврик для мыши CACTUS CS	Ноутбук ASUS Zenbook UM43
8	5	0,17	0,50	2,00	Ноутбук ASUS VivoBook E40	Гарнитура игровая OKLICK

Рис. 2. Таблица ассоциативных правил в Loginom

Model | Settings | Summary | Annotations

Sort by: Confidence % 26

Consequent	Antecedent	Rule ID	Instances	Support %	Confiden...	Rule Sup...	Lift	Deploya
Bakery goods	Milk Frozen foods	5	85	10.814	83.529	9.033	1.948	1.781
Bakery goods	Alcohol Tinned Goods Ready made	18	95	12.087	83.158	10.051	1.94	2.036
Bakery goods	Frozen foods Tinned Goods Snacks	22	90	11.45	82.222	9.415	1.918	2.036
Ready made	Alcohol Tinned Goods Bakery goods	17	97	12.341	81.443	10.051	1.654	2.29
Ready made	Alcohol Tinned Goods Snacks	19	91	11.578	79.121	9.16	1.607	2.417

Рис. 3. Таблица ассоциативных правил в IBM SPSS Modeler

В системе IBM SPSS Modeler кроме Таблицы правил имеется дополнительная встроенная визуализация результатов с помощью специализированных графиков Web и Directed Web.

Диаграммы Web (рис. 4) интерпретируют силу взаимосвязи между значениями двух или более символических полей изменением типов линий. Узел Web можно использовать, например, для исследования взаимосвязи покупок различных товаров.

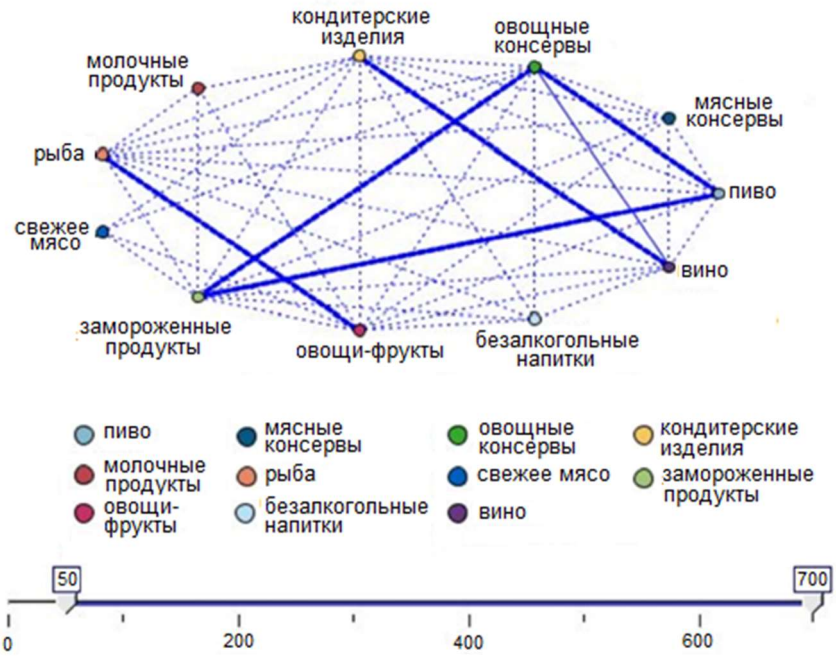


Рис. 4. График Web, показывающий взаимосвязь между покупками товаров

Диаграммы Directed Web показывают значения полей *Условие* с единственным полем *Следствие*. Эти связи однонаправленные, то есть они определены только в одну сторону, это хорошо видно на рис. 5.

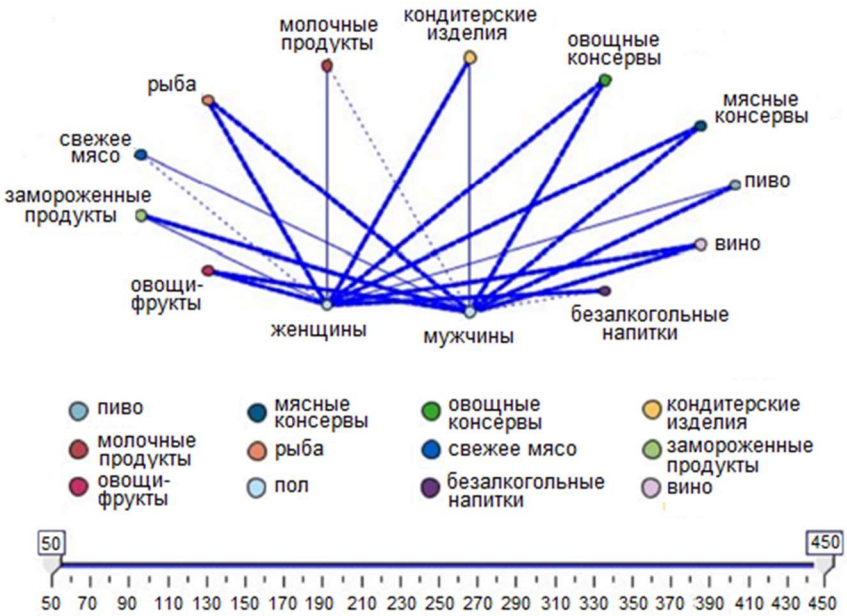
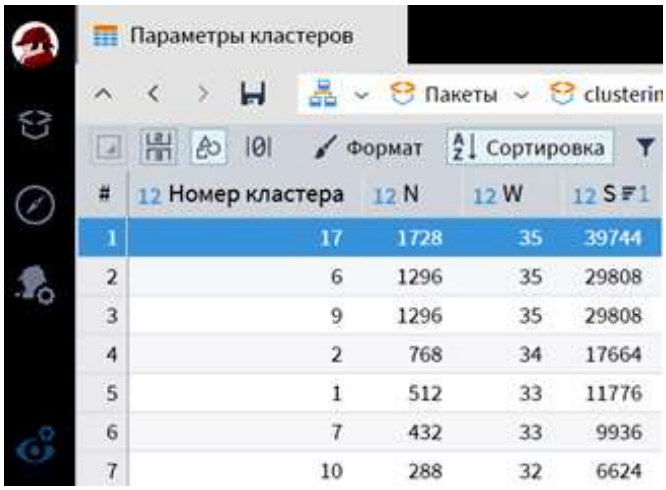


Рис. 5. График Directed Web, показывающий взаимосвязь между покупкой товаров и полом покупателя

Диагностическая аналитика. Кластерный анализ

Таблицы кластеров в рассматриваемых системах отражают разбиение входного набора на кластеры и позволяют оценить вероятность принадлежности объекта к определенной группе.

Таблица в Loginom отражает информацию о параметрах разбиения и позволяет оценить кластеры по различным характеристикам: количество транзакций (N), ширина кластера (W), мощность кластера (S). В таблице возможны сортировка, фильтрация, форматирование, детализация, как видно на рис. 6.



#	Номер кластера	N	W	S
1	17	1728	35	39744
2	6	1296	35	29808
3	9	1296	35	29808
4	2	768	34	17664
5	1	512	33	11776
6	7	432	33	9936
7	10	288	32	6624

Рис. 6. Таблица кластеров в Loginom

В *Таблице кластеров* 1С вычисляются центры кластеров, расстояние между кластерами по одной из метрик, количество и процент объектов кластера (рис. 7).

Результат:

Кластеры

№ кластера	Количество объектов	Процент
1	1	33,33
2	1	33,33
3	1	33,33

Центры кластеров

Расстояния между кластерами

Кластеры	1	2	3
1		17,1109	10,7424
2	17,1109		10,1467
3	10,7424	10,1467	

Рис. 7. Таблица кластеров в 1С

Таблица кластеров в IBM SPSS Modeler (рис. 8) отличается от таблицы 1С наличием графического визуализатора пропорций кластера.

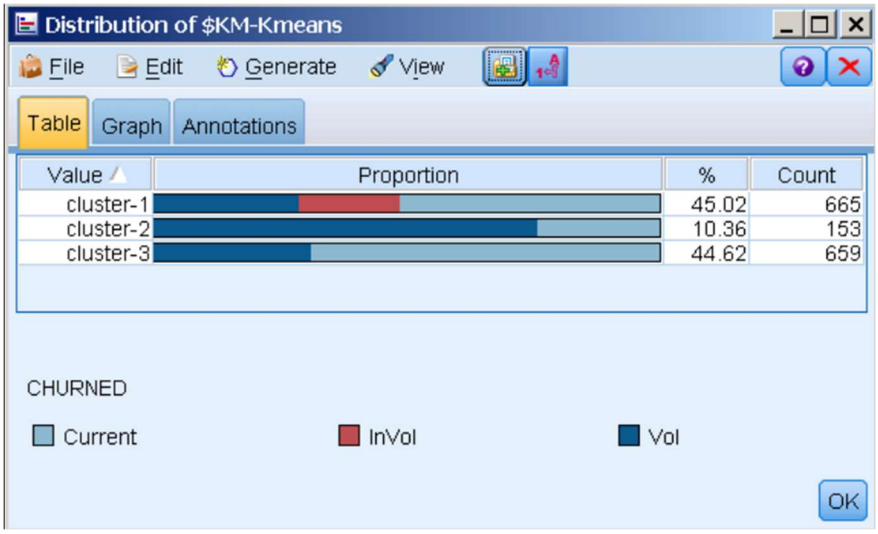
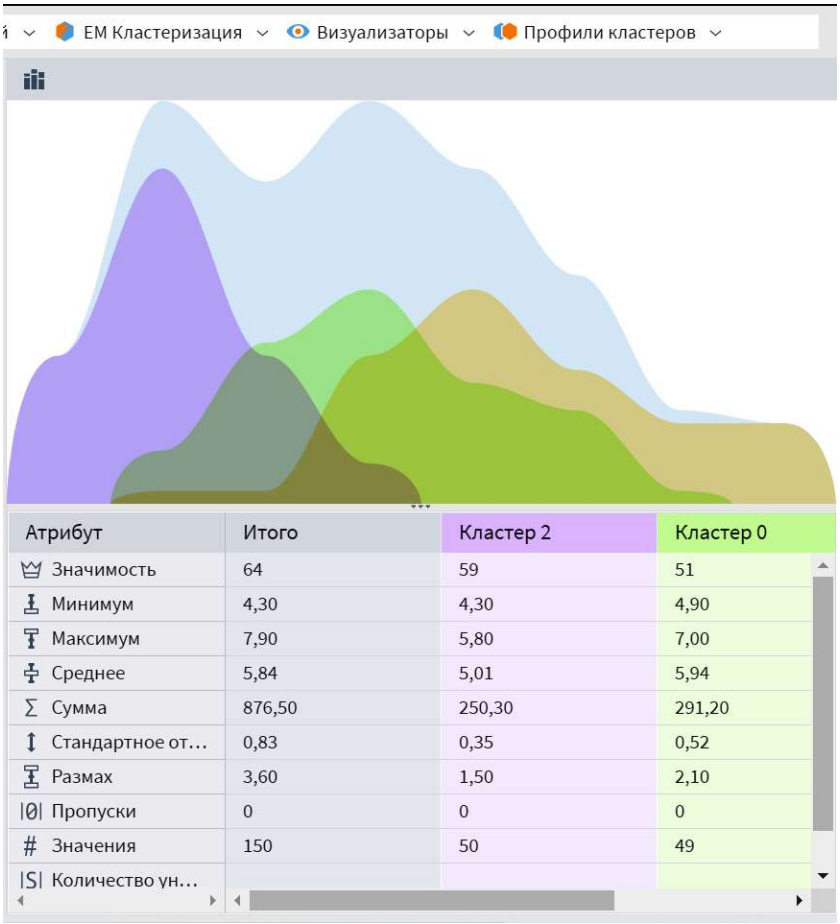


Рис. 8. Таблица кластеров в IBM SPSS Modeler

Диагностическая аналитика кроме описания статистики кластеров требует дополнительных визуальных представлений, так как должна выявить факторы, оказывающие влияние на целевые параметры.

В системе Loginom визуализатор *Профили кластеров* предназначен для просмотра различных статистических показателей кластеров, просмотра структуры кластеров и сравнения их между собой (рис. 9).



Атрибут	Итого	Кластер 2	Кластер 0
Значимость	64	59	51
Минимум	4,30	4,30	4,90
Максимум	7,90	5,80	7,00
Среднее	5,84	5,01	5,94
Сумма	876,50	250,30	291,20
Стандартное от…	0,83	0,35	0,52
Размах	3,60	1,50	2,10
Пропуски	0	0	0
Значения	150	50	49
Количество ун…			

Рис. 9. Инструмент визуализации «Профили кластеров» в Loginom

При выводе данных кластерного анализа в 1С: Предприятие, если используется алгоритм, отличный от алгоритма k-средних, результаты выводятся в виде дендрограммы. *Таблицы профилей кластеров* описывают характеристики найденных кластеров (их количество, центры, расстояния между ними), но не показывают, какие объекты (на рис. 10 контрагенты компании) в какие кластеры входят. *Дендрограмма*, как специализированный визуализатор, показывает, насколько одни объекты похожи на другие объекты или на группы объектов.

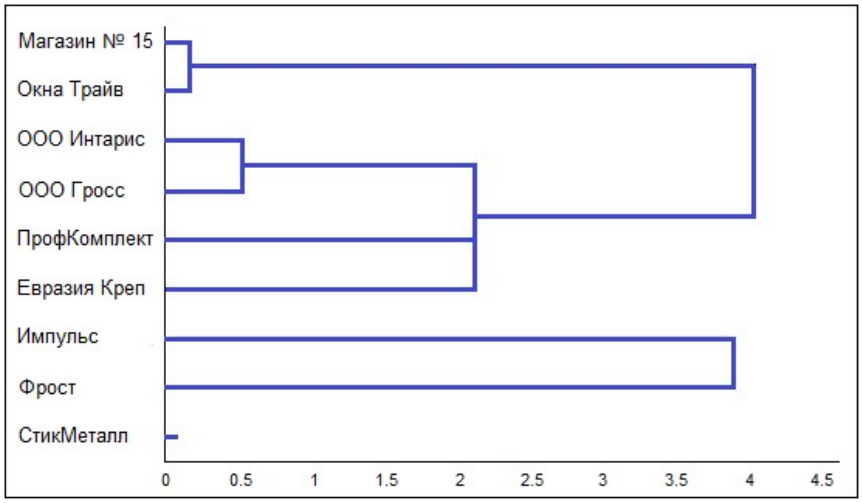


Рис. 10. Дендрограмма для визуализации распределения анализируемых объектов по кластерам в 1С

В системе IBM SPSS Modeler категориальные показатели выводятся в виде точечных диаграмм, как это показано на рис. 11.

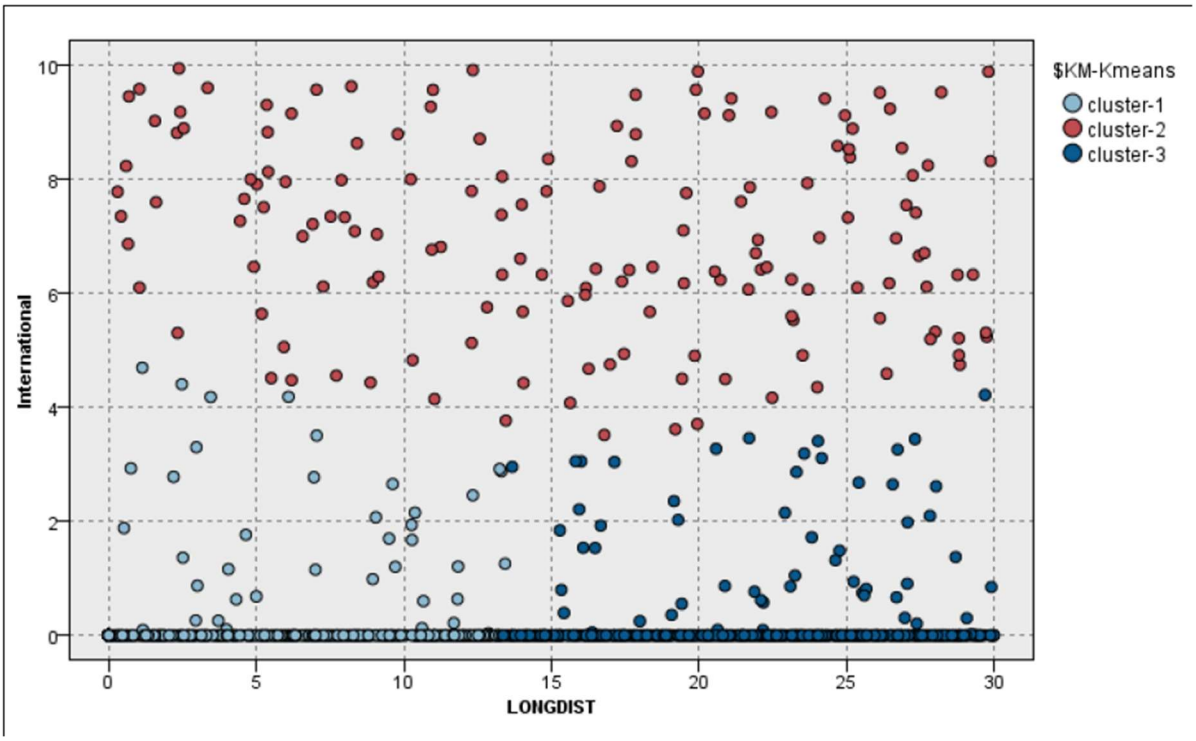


Рис. 11. Визуализация принадлежности объектов кластерам в IBM SPSS Modeler

Предикативная аналитика. Классификация и регрессия

Разбиение множества объектов или наблюдений на априорно заданные группы, называемые классами, внутри каждой из которых они предполагаются похожими друг на друга, имеющими примерно одинаковые свойства и признаки, является одной из важнейших задач Data Mining.

В исследуемых системах классификация реализуется специализированными обработчиками, основанными или на статистических методах: логистическая регрессия, дискриминантный анализ, или на методах машинного обучения: нейронные сети, деревья решений, метод k-ближайших соседей, машины опорных векторов и др. Платформа 1С реализует статистические методы, Loginom и IBM SPSS, являясь аналитическими системами, реализуют оба вида обработчиков. В целях выравнивания сравниваемых возможностей были рассмотрены обработчики, использующие статистические методы.

Преимуществом статистических методов является их хорошая математическая обоснованность, недостатком — низкая объясняющая способность. Использование вероятностных оценок позволяет с высокой точностью предсказать, к какому классу относится объект, но не позволяет сказать почему. Поэтому результаты статистических методов классификации могут оказаться сложными для понимания и интерпретации результатов.

В системе 1С визуализатором результатов прогнозной аналитики является *Дерево решений* (рис. 12).

Анализ данных | Прогнозирование

Общая статистика | Поиск ассоциаций | Поиск последовательностей | Дерево решений | Кластеризация

Результат:

Дерево решений

Условие	Решение	Вероятность	Количество
Корень	СостояниеВзаимоотношений = Не прекращены	66,67	3
КоличествоРозничныхТочек <= 9	**СостояниеВзаимоотношений = Не прекращены**	**100,00**	**2**
КоличествоРозничныхТочек > 9	**СостояниеВзаимоотношений = Прекращение контрагентом**	**100,00**	**1**

Рис.12. Дерево решений в системе 1С

В аналитической платформе Loginom для визуализации статистических методов классификации и регрессии используются *Таблицы* и *Деревья регрессий* (рис.13).

рессия | Логистическая регрессия | Сценарий | Логистическая регрессия | Визуализаторы

Дерево | Нулевые значения

Атрибут	Коэффициент	Стандартная ошибка	Статистика Вальда	Р-значение
Константа	-2,105737	0,313109	45,229049	1,752853e-11
Тип планировки				
брежневка	0,114181	0,176319	0,419361	0,517256
Нестандартная	0,408625	0,183004	4,985708	0,025558
свердловский ва...	-0,232423	0,299240	0,603280	0,437329
сталинка	0,192056	0,153621	1,562974	0,211230
хрущевка	0,325572	0,196280	2,751328	0,097174
Количество комнат	0,468378	0,156747	8,928868	0,002807
Первый/Последний ...				
Истина	0,253268	0,104057	5,924036	0,014936
Общая площадь (м2)	-0,021558	0,010676	4,077524	0,043457
Площадь кухни (м2)	0,087484	0,039383	4,934441	0,026327
Стоимость (т.руб.)	0,000241	0,000239	1,015788	0,313520

Рис. 13. Дерево логистической регрессии в Loginom

В SPSS визуализацию результатов прогнозирования можно увидеть с помощью *Таблиц*, *Деревьев классификации* (рис.14) и *Деревьев регрессии*.

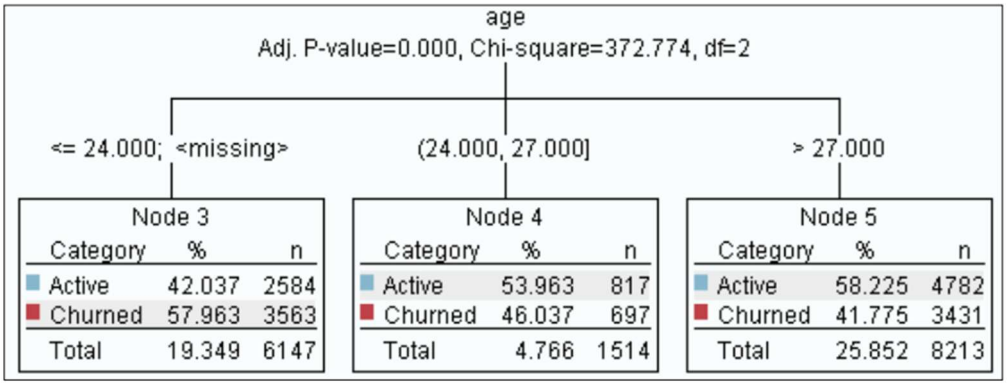


Рис. 14. Узел CHAID в IBM SPSS генерирует деревья решений, используя статистику хи-квадрат для определения оптимальных расщеплений

Визуализация метрик моделирования

Оценка точности моделей выполняется специализированными визуализаторами. В табл. 4 приведен некоторый набор визуализаторов оценки качества для Data Mining.

Алгоритмы поиска ассоциативных правил позволяют находить закономерности между связанными событиями. Для оценки значимости правил используются объективные и субъективные показатели. Объективными являются такие метрики, как поддержка и достоверность правил, они могут применяться независимо от конкретной задачи. Субъективные метрики связаны с контекстом решаемой задачи, к ним относятся: лифт (интерес), левередж (плечо, рычаг), улучшение (improvement).

Объективные метрики вычисляются во всех аналитических приложениях, субъективные – только в интерактивных системах, т. е. системах, в которых пользователь имеет возможность вводить дополнительную информацию, определяемую контекстом задачи.

Таблица 4 – Задачи аналитики и специализированные визуализаторы метрик модели

Задачи аналитики	Инструменты оценки качества моделирования
Ассоциативные правила	Таблица - численные метрики для поддержки, достоверности и значимости правил
Кластерный анализ	Сравнительная гистограмма кластеров
Классификация Регрессионный анализ	Панель информации о модели Диаграммы оценки качества классификации Матрица ошибок классификации с численными метриками качества классификации

Основными метриками кластерного анализа, исходя из целей кластеризации, являются расстояния: внутри кластера расстояние должно быть максимально близким, между разными группами – наибольшим. Для оценки полученных расстояний используются специализированные панели сравнения кластеров. Например, в системе Loginom панель сравнения включает сравнительную гистограмму кластеров (совмещение гистограммы кластеров и генеральной совокупности) и сводную статистику кластеров (рис.15).

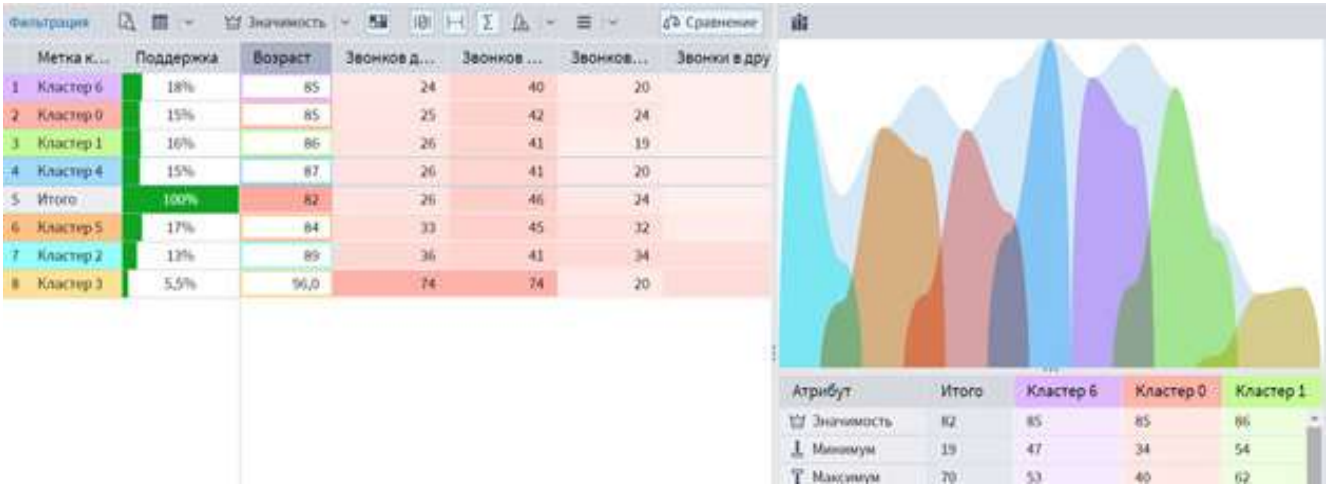


Рис. 15. Панели оценки качества кластеризации в Loginom

Как Loginom, так и IBM SPSS располагают большой линейкой диаграмм для визуализации качества регрессии и классификации. Например, диаграмма оценки качества классификации ROC-кривая (Receiver Operator Characteristic) показывает зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров (рис. 16), диаграмма равновесия показывает точность прогноза (рис. 17).

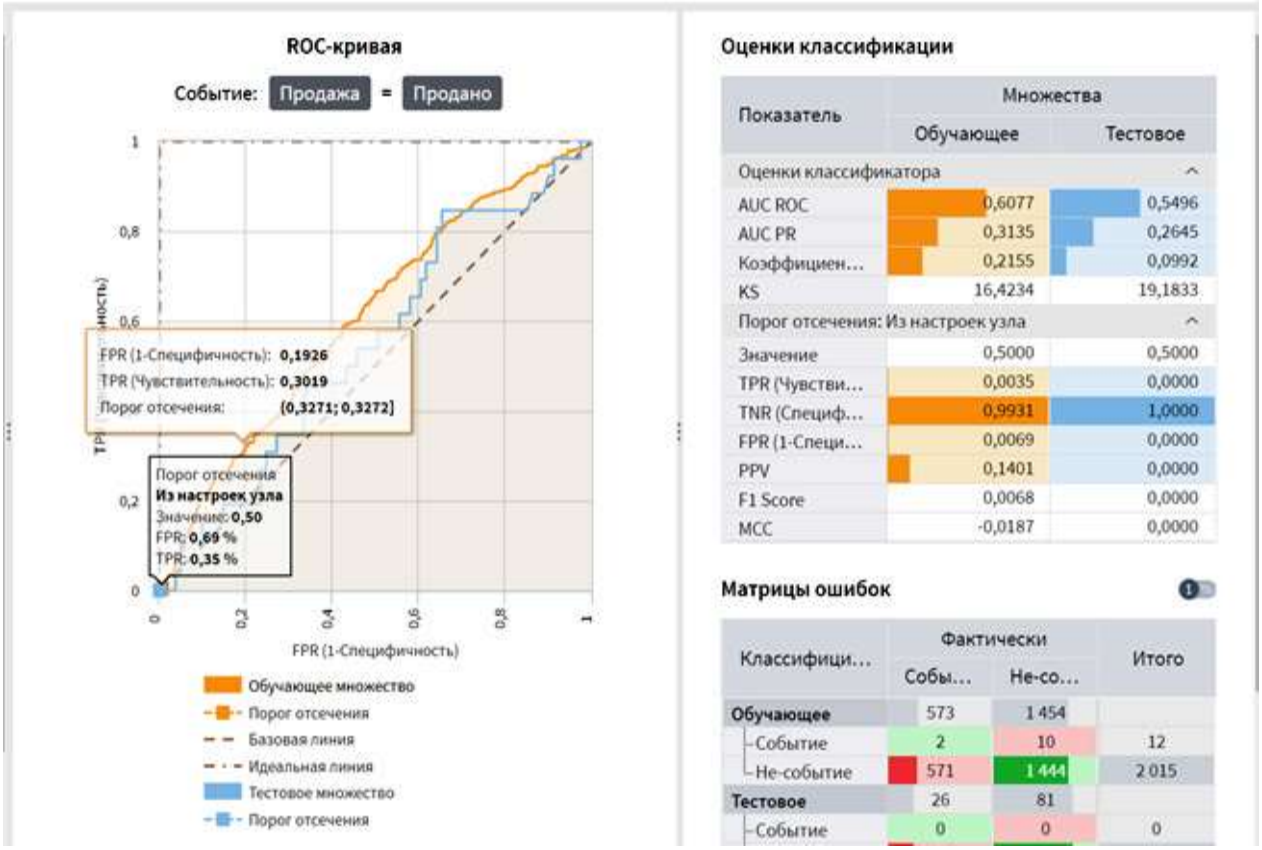


Рис. 16. Диаграмма оценки качества классификации в Loginom – ROC-кривая

Рис. 17. Диаграмма оценки качества классификации в Loginom – Диаграмма равновесия

В системе 1С качество классификации визуализируется матрицей ошибок с численными метриками качества классификации.

Количество и разнообразие инструментов визуализации подтверждают особую роль визуализации данных в системах, где данные становятся источником новых знаний и основой поддержки принятия решений. Выполненный анализ показывает, что базовые инструменты визуализации: *Таблица, Диаграмма, График, Деревья* присутствуют во всех рассмотренных системах. Разнообразие инструментария определяется специализацией системы, уровнем интерактивности интерфейса и классом решаемых задач.

Тезис о недостаточной поддержке «простых» моделей средствами визуализации подтвердился для систем Loginom и 1С. Результаты моделирования ассоциативных правил в этих системах интерпретируются исключительно *Таблицей* (рис.1 и рис.2), оценка качества модели выполняется также табличным представлением трех метрик (поддержка, достоверность, лифт). В то же время в системе IBM SPSS представлена дополнительная визуализация для правил (рис.3 и рис.4), которая имеет два важных достоинства: наглядность и выделение интенсивностью цвета «сильных» правил, имеющих максимальные показатели при их оценке.

Заключение

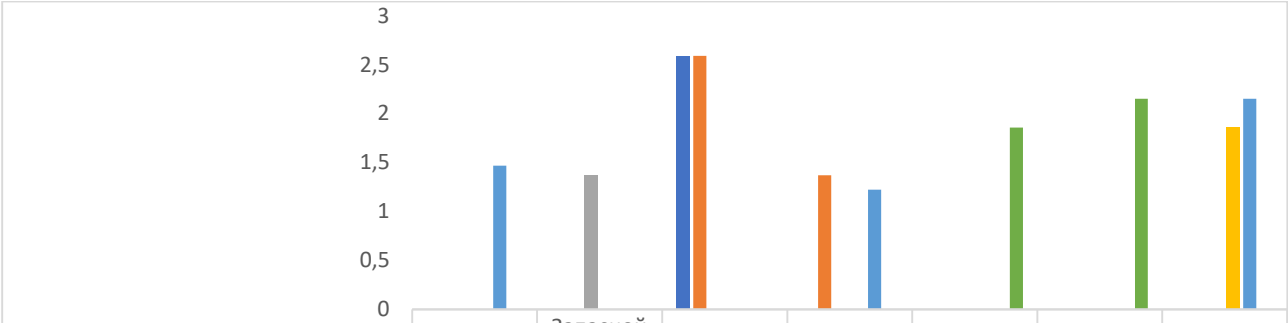
Авторы статьи анализируют средства визуализации результатов Data Mining для применения в учебном процессе. Ранее в статье [8] были рассмотрены визуализаторы общего назначения, эффективность которых при изучении сложных методов анализа становится недостаточной. В предложенном анализе рассматриваются специализированные инструменты визуализации для интерпретации результатов в трех аналитических пакетах: Loginom, IBM SPSS Modeler, 1С: Предприятие. Выбор систем объясняется использованием перечисленного программного обеспечения в дисциплинах образовательной программы.

В качестве основы для анализа средств визуализации Data Mining авторы использовали классификацию визуализаторов в разрезе используемых аналитических методов и категорий практических задач. Таким образом, все специализированные визуализаторы были сгруппированы для задач ассоциативных правил (дескриптивная аналитика), кластеризации (диагностическая аналитика), задач классификации и регрессионного анализа (предикативная аналитика).

В исследовании были учтены особенности специализации программных систем: пакеты IBM SPSS и Loginom являются аналитическими платформами, в то время как платформа 1С: Предприятие – система автоматизации предприятия, поддерживающая встроенную аналитику данных [11]. При сравнении компонентов визуализации учитывались методы, на которых реализованы компоненты Data Mining, например, для моделей классификации рассматривались только компоненты, поддерживающие статистические методы.

Анализ возможностей встроенных специализированных визуализаторов в аналитических системах различного класса ПО подтвердил априорное авторское представление: с усложнением моделей от описательной к предсказательной аналитике увеличивается количество и сложность имеющихся средств интерпретации и, в частности, средств визуализации результатов. Если для предикативной аналитики в любой из рассмотренных систем имеются эффективные встроенные средства визуализации, то для условно «простых» моделей анализа, например ассоциативных правил, такие визуализаторы представлены не во всех системах.

В системах 1С и Loginom было обнаружено отсутствие визуальных представлений для ассоциативных правил, *Таблица правил* – единственное средство представления результатов анализа. Пакет IBM SPSS для ассоциативных правил имеет достаточно наглядные диаграммы. В целях расширения возможностей визуализации Loginom и 1С авторы предлагают использовать, например, инструмент *Сводная диаграмма* офисного пакета Excel (рис.18).



	Антистатик спрей	Запасной баллон для освежителя	Освежитель воздуха	Пена/соль для ванн	Салфетки бумажные	Средство для мытья посуды	Средство для чистки плит
Бумага туалетная			2,59				
Запасной баллон для освежителя			2,59	1,37			
Пена/соль для ванн		1,37					
Салфетки бумажные							1,86
Средство для мытья посуды	1,47			1,22			2,15
Средство для чистки плит					1,86	2,15	

Рис. 18. Сводная диаграмма в Excel для визуализации ассоциативных правил из Loginom или 1С

Еще одним средством визуализации ассоциативных правил, поддерживаемым Excel, является *Тепловая карта*, в которой путем использования иерархической кодировки цвета достигается фактически трехмерное отображение данных [12]. Тепловая карта, пример

которой приведен на рис.19, обеспечивает удобство восприятия представленной информации и позволяет оценить взаимосвязи между отдельными событиями.

Названия строк	Антистатик спрей	Запасной баллон для освежителя	Освежитель воздуха	Пена/соль для ванн	Салфетки бумажные	Средство для мытья посуды	Средство для чистки плит
Бумага туалетная			2,59				
Запасной баллон для освежителя			2,59	1,37			
Пена/соль для ванн		1,37					
Салфетки бумажные							1,86
Средство для мытья посуды	1,47			1,22			2,15
Средство для чистки плит					1,86	2,15	

Рис. 19. Тепловая карта в Excel для визуализации ассоциативных правил

Платформа 1С: Предприятие предоставляет разработчику интегрированный набор инструментов, необходимых для быстрой разработки, распространения и поддержки прикладного решения автоматизации бизнеса. При этом платформа имеет встроенные компоненты Data Mining как для описательной аналитики, так и для прогнозирования событий. Анализ показал, что разнообразие и функциональность средств визуализации в 1С уступает специализированным аналитическим пакетам, однако эффект достигается благодаря общему набору средств и их тесной интеграции.

Практическая значимость результатов приведенного в статье анализа современных средств визуализации не ограничивается рамками образовательного процесса в вузе, так как подготовка выпускников направлена на формирование специалистов, способных решать практические задачи в прикладных областях, связанных с обработкой больших данных, решением аналитических задач и потому заинтересованных в использовании наиболее эффективных средств визуализации результатов Data Mining. Одним из примеров может служить организация транспортного обслуживания, в частности, прогнозное техническое обслуживании авиакомпаний, являющееся весьма актуальной проблемой современной авиационной индустрии [13].

СПИСОК ЛИТЕРАТУРЫ

1. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ. – М.: Мир, 1981. – 696 с.
2. Emerson J.W., Green W.A., Schloerke B., Crowley J., Cook D., Hofmann H., Wickham H. The Generalized Pairs Plot // Journal of Computational and Graphical Statistics. – 2013. – Vol. 22, N 1. – P. 79-91. doi: 10.1080/10618600.2012.694762.
3. Романова И. К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах // Наука и

Образование. МГТУ им. Н. Э. Баумана. Электрон. журн. 2016. № 3. С. 133–167. doi: 10.7463/0316.0834876.
4. Пятецкий-Шапиро Г. Data Mining и перегрузка информацией // Вступительная статья к книге: Анализ данных и процессов / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. 3-е изд. перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
5. Что такое аналитика данных? [Электронный ресурс] // Intel – URL: https://www.intel.ru/content/www/ru/ru/analytics/what-is-data-analytics.html (дата обращения: 21.02.2022).
6. 10 инструментов аналитики данных. [Электронный ресурс] // АСУ-АНАЛИТИКА системы принятия решений – URL: https://asu-analitika.ru/10-instrumentov-analitiki-dannyh/ (дата обращения: 21.02.2022).
7. Предикативная аналитика URL: https://www.tadviser.ru/a/180866 (дата обращения 21.02.2022).
8. Богословская Н.В., Бржезовский А.В. Визуализация для интерпретации данных при использовании аналитических пакетов // Обработка, передача и защита информации в компьютерных системах: Материалы I Всерос. науч. конф. – СПб., 2020. – С. 277-281.
9. LOGINOM. Аналитическая платформа Loginom [Электронный ресурс]. – URL: https://loginom.ru/ (дата обращения 21.02.2022).
10. IMB SPSS Statistics [Электронный ресурс]. – URL: https://www.ibm.com/ru-ru/products/spss-statistics (дата обращения 21.02.2022).
11. Фирма «1С» [Электронный ресурс]. – URL: https://1c.ru/ (дата обращения 21.02.2022).
12. Wilkinson L., Friendly M. The History of the Cluster Heat Map // American Statistician. – 2009. – Vol. 63, N 2. – P. 179–184.
13. Richards J., Seidakhmetov B.K., Zhunussova M.A. Digital Transformation in the Aviation Industry//Вестник академии гражданской авиации. – 2020. – № 2(17). – С. 23-27

ИНФОРМАЦИЯ ОБ АВТОРАХ

Богословская Наталья Валентиновна –
к.т.н., доцент кафедры информационно-сетевых технологий
Санкт-Петербургский государственный университет аэрокосмического приборостроения
190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А
E-mail: nvbogoslov@mail.ru

Бржезовский Александр Викторович –
к.т.н., доцент кафедры информационно-сетевых технологий
Санкт-Петербургский государственный университет аэрокосмического приборостроения
190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А
E-mail: avb@aanet.ru

Красильникова Ольга Ивановна –
к.т.н., доцент кафедры информационно-сетевых технологий
Санкт-Петербургский государственный университет аэрокосмического приборостроения
190000, Санкт-Петербург, ул. Большая Морская, д. 67, лит. А
E-mail: oikrasilnikova@yandex.ru

INFORMATION ABOUT THE AUTHORS

Bogoslovskaia Natalia Valentinovna –
PhD, associate Professor
Saint-Peterdburg State University of Aerospace Instrumentation
67, BolshayaMorskaia str.Saint-Petersburg,190000, Russia
E-mail: nvbogoslov@mail.ru

Brzhezovskii Aleksandr Viktorovich –
PhD, associate Professor
Saint-Peterdburg State University of Aerospace Instrumentation
67, BolshayaMorskaia str.Saint-Petersburg,190000, Russia
E-mail: avb@aanet.ru

Krasilnikova Olga Ivanovna –
PhD, associate Professor
Saint-Peterdburg State University of Aerospace Instrumentation
67, BolshayaMorskaia str.Saint-Petersburg,190000, Russia
E-mail: oikrasilnikova@yandex.ru