

БОДЯКИНА Е., ст. гр. МЭД 08
Науч. руков.: Овсянников В.П., к.т.н., доц.
ГВУЗ "Донецкий национальный технический университет",
г. Донецк

ПРИМЕНЕНИЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ DEDUCTOR ДЛЯ АНАЛИЗА ДАННЫХ, ХАРАКТЕРИЗУЮЩИХ РАБОТУ УГОЛЬНОГО ПРЕДПРИЯТИЯ

Приведены основные сведения о функциональных возможностях аналитической платформы Deductor. Рассмотрен пример использования этой системы для анализа данных, характеризующих работу угольного предприятия.

Актуальность. Deductor [1, 2] является аналитической платформой - предназначенной для создания законченных прикладных решений в области анализа данных. Реализованные в Deductor технологии позволяют на базе единой архитектуры пройти все этапы построения аналитической системы от создания хранилища данных до автоматического подбора моделей и визуализации полученных результатов.

Цель исследования. Привести основные сведения о функциональных возможностях аналитической платформы Deductor. Показать эффективность этой системы на примере анализа данных, характеризующих работу угольного предприятия.

Основная часть. Deductor состоит из пяти частей:

Deductor Warehouse - многомерное хранилище данных, аккумулирующее всю необходимую для анализа предметной области информацию. Использование единого хранилища позволяет обеспечить непротиворечивость данных и централизованное хранение, а также автоматически обеспечивает всю необходимую поддержку процесса анализа данных.

Deductor Studio - программа, реализующая функции импорта, обработки, визуализации и экспорта данных. Deductor Studio может функционировать и без хранилища данных, получая информацию из любых других источников, но наиболее оптимальным является их совместное использование. В Deductor Studio включен полный набор механизмов, позволяющий получить информацию из произвольного источника данных, провести весь цикл обработки (очистку, трансформацию данных, построение моделей), отобразить полученные результаты наиболее удобным образом (OLAP, таблицы, диаграммы, деревья и т.д.) и экспортировать их в наиболее распространенные форматы.

Deductor Viewer - программа, ориентированная на конечного пользователя и предназначенная для просмотра подготовленных при помощи Deductor Studio отчетов.

Deductor предназначена для решения широкого спектра задач,

связанных с обработкой структурированных и представленных в виде таблиц данных. При этом область приложения системы может быть практически любой - механизмы, реализованные в системе, с успехом применяются на финансовых рынках, в страховании, торговле, телекоммуникациях, промышленности, медицине, в логистических и маркетинговых задачах и множестве других.

Кратко рассмотрим наиболее распространенные классы задачи, на решение которых рассчитан Deductor:

Системы отчетности. В Deductor включено многомерное хранилище данных Deductor Warehouse, позволяющее быстро создавать системы корпоративной отчетности - финансовой, аналитической, маркетинговой и т.д. Хранилище данных решает задачи консолидации, обеспечения целостности и непротиворечивости данных. Оно оптимизировано для обработки больших объемов данных, поэтому при работе с системой задержки, связанные с загрузкой и сохранением информации, сведены к минимуму.

Благодаря наличию большого набора визуализаторов хранящиеся в Deductor Warehouse данные можно просматривать различными способами: в виде OLAP кубов, таблиц, диаграмм, гистограмм и т.д.

Data Mining проекты. Data Mining - это процесс извлечения из исходных данных ранее неизвестных, нетривиальных и практически ценных зависимостей (знаний). Сфера применения Data Mining не ограничена одной прикладной областью. Он может применяться везде, где возникает потребность в глубоком анализе данных, хотя чаще всего речь идет об анализе коммерческой информации. Среди множества задачи, решаемых при помощи методов Data Mining, можно выделить такие:

- Анализ и управление рисками;

- Оценка кредитоспособности физических и юридических лиц;

- Определение профилей клиентов, т.е. наиболее характерных особенностей их поведения;

- Промышленная диагностика, обнаружение источников и причин возникновения дефектов;

- Идентификация критических ситуаций;

- Стимулирование продаж;

- Сегментация клиентов, продуктов, услуг;

- Оценка факторов, влияющих на лояльность клиентов и множество других...

Очистка данных. Очистка данных даже без последующей обработки сама по себе представляет значительный интерес. На практике исходные данные чаще всего бывают плохого качества. Это связано с объективными причинами, и избавиться от этого, применяя какие-либо административные или организационные методы, невозможно

Прогнозирование. Прогнозирование - одна из наиболее востребованных задач анализа. В Deductor включено несколько механизмов построения прогностических моделей, в том числе с использованием

самообучающихся алгоритмов. Но кроме этого в системе имеются механизмы очистки данных, которые приносят особенно большую пользу при решении задачи прогнозирования. Имеющийся набор позволяет получать качественные прогнозы и обладает большими возможностями по адаптации, т.е. способен подстраиваться под изменяемую ситуацию.

Моделирование. Построение моделей - наиболее универсальный способ анализа. В действительности при анализе мы в том или ином виде всегда имеем некую модель исследуемого процесса или объекта. Но не всегда эта модель формализована, т.е. описана таким образом, чтобы ей мог воспользоваться кто-то в режиме "черного ящика": подали данные на вход, а на выходе получили результат.

Анализ "что-если". При помощи Deductor можно не только строить модели, но и провести анализ по принципу "что-если", т.е. оценить, как может измениться тот или иной показатель при изменении любого влияющего фактора. Для реализации этого простого в использовании и одновременно мощного механизма предназначен специальный визуализатор. При этом не имеет значения, каким способом производилось построение модели, работа со всеми алгоритмами выполняется одинаково. Результаты можно просмотреть как в табличном, так и в графическом виде. Такого рода механизмы анализа являются готовым инструментом для оптимизации экономических процессов.

Рассмотрим пример использования Deductor для анализа данных, характеризующих работу угольного предприятия.

Данные [3] которые изначально представлены в Excel были конвертированы в txt-формат а затем успешно экспортированы в платформу Deductor Academic рис. 1. Кстати, можно также отметить, что применение аналитической платформы Deductor Academic не требует получения какой-либо лицензии.

дата	добыча угля	добыча неэнергетического угля	добыча коксующегося угля	зольность	производственные мощности	использование мощностей	численность работающих	численность работающих по добыче
1990	150	85	65	28	200	72	905	540
1991	135,6	80	55,6	29,8	190	70	870	510
1992	122	73	49	30	180	68	825	490
1993	110	65	45	31	165	64	790	470
1994	95	58	37	32	150	61	750	438
1995	81	50	31	33	140	59	710	420
1996	74,8	45	29,8	34	130	58	680	400
1997	73	41	32	35	120	60	625	370
1998	80	40	40	35,5	112	66	600	350
1999	88	41	47	36	110	70,5	560	314
2000	89,3	42	47,3	36,5	110,3	72	530	300
2001	89	42	47	37	110,7	71	480	280
2002	88	43	45	37,5	114	69	445	250
2003	85	44	41	37,8	120	67	400	230
2004	82	45	37	37,5	122	64	354	213
2005	80	46	34	38,1	128	61	310	200
2006	78	47	31	38,5	130	58	280	180
2007	75	48	27	39	134	56	230	165
2008	72	50	22	39,5	136	54	180	150
2009	69	51	18	40	138	52	130	135

Рисунок 1. Данные, характеризующие работу угольного предприятия в

аналитической платформе Deductor Academic

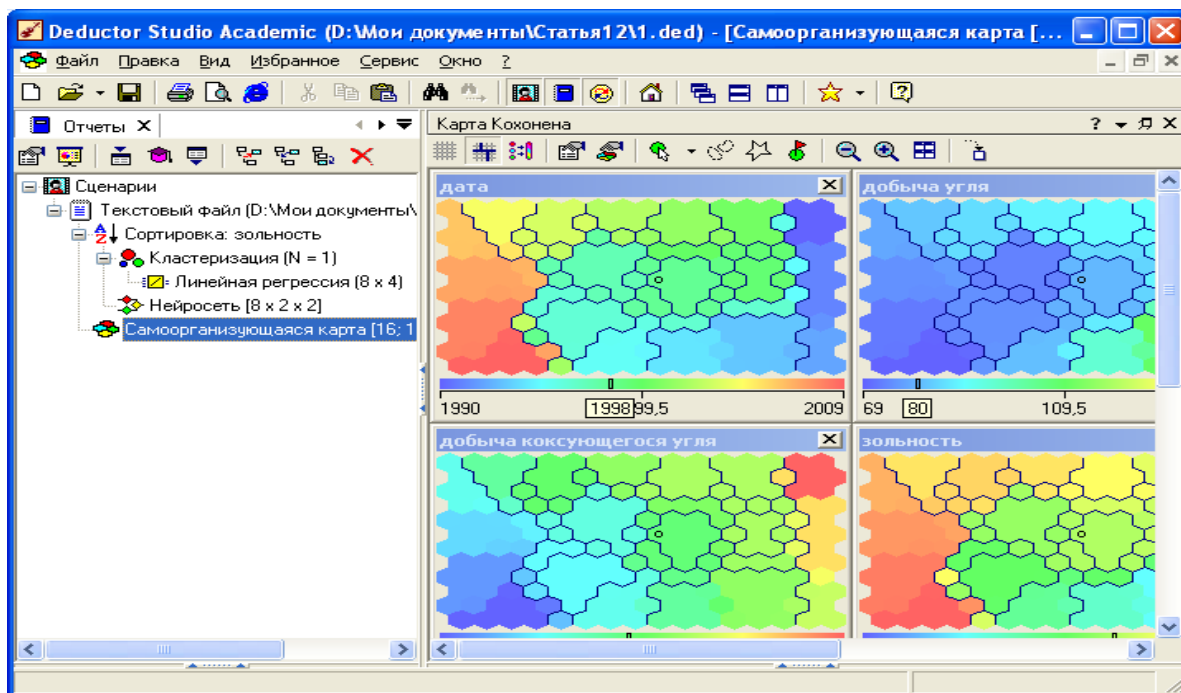


Рисунок 2. Пример анализа данные, характеризующие работу угольного предприятия в аналитической платформе Deductor Academic

Дальнейший анализ, по сути, становится «делом техники» и легко осуществляется в рамках алгоритмов анализа, которые предоставляются аналитической платформой Deductor (рис. 2).

Можно, например, выполнить факторный анализ (рис. 3) и показать, что работа угольного предприятия характеризуется тремя значимыми параметрами, что, кстати, совпадает с результатом, представленным в [1] причем, этот результат достигается без дополнительных усилий и без использования дорогостоящих лицензионных программных комплексов.

Фактор_1	Фактор_2	Фактор_3
2980558614331	0,473748796015799	-1,18770976065774
2753233232498	0,673803411648255	-0,729162277252817
2270329926695	0,748044022281213	-0,201650061048292
5162523407083	0,784578348958012	0,353040371456205
2483532898479	0,900968758804714	0,95252037758392E
5713549275159	0,887818859391975	1,46398969164684
6347231148018	0,656169064983419	1,6329151847205E
4730257758844	-0,033855044360295	1,4485841324674E
2869954454631	-1,28794949528158	0,83999652574466E
1252534892537	-2,12524164274696	0,12625929651842E
7098109864922	-2,27792772152025	-0,15422918626154E
2777887832752	-2,13395266643204	-0,32332800050971E
8395528777738	-1,71191350568463	-0,45779687793315E
3,101651653679	-1,13880190083036	-0,51454612543890E
5032480036456	-0,508769010817849	-0,45630224173667E
7778828617925	0,133354148775632	-0,51314092730908E
6124182611991	0,69267337579813	-0,51845712394735E

Рисунок 3. Факторный анализ данные, характеризующие работу угольного предприятия в аналитической платформе Deductor Academic

Библиографический список

1. Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям (+ CD): учеб. пособие.— 2-е изд., перераб. и доп.— СПб.: Питер, 2010. — 704 с.
2. Кацко И.А., Паклин Н.Б. Практикум по анализу данных на компьютере: Учеб. пособие для вузов - М.: Издательство "Колос", 2009. - 278 с.; 60x88/16. ISBN 978-5-9532-0624-2 (в пер.), 1000 экз. УДК 681.5(075.8) ББК 32.81я73 ГРИФ "Допущено УМО по образованию в области прикладной информатики".
3. Сервуля Ф. Овсянников В.П., Алгоритм анализа данных характеризующих работу угольного предприятия средствами программного комплекса SPSS Сучасні проблеми управління виробництвом: Тези доповідей IV Міжнар.н.-пр. конф., м. Донецьк, ДонНТУ, 14–15 жовтня 2010 р.

ГОРБАЧ Н.М., ст. гр. МО-07а
 Наук. керівн.: Лихошерст Є.В., асист.
 ДВНЗ "ДонНТУ" Автомобільно-дорожній інститут,
 м. Горлівка

УДОСКОНАЛЕННЯ СИСТЕМИ ОЦІНКИ ПОТЕНЦІАЛУ ПІДПРИЄМСТВА