

Краткое резюме некоторых прикладных проектов Рыжова А.П.

1. Системы оценки и мониторинга сложных процессов. Задача оценки и мониторинга включает в себя оценку состояния некоторого процесса и моделирование возможных вариантов его развития на базе всей доступной информации (структурированной, слабоструктурированной, неструктурированной). Такие задачи решаются многими специалистами из различных областей управления экономическими, социальными и политическими процессами. Разработанная технология оценки и мониторинга позволяет делать это максимально эффективно в рамках человеко-машинных систем. К настоящему времени разработаны:

- Система мониторинга и оценки мирной ядерной деятельности стран в интересах управления обеспечения международных гарантий МАГАТЭ (Development of an Intelligent System for Monitoring and Evaluation of Peaceful Nuclear Activity, IAEA, Vienna)
- Система «Мониторинг - 1» (НИЦ «Контур» ФАПСИ РФ)
- Система оценки и мониторинга риска атеросклеротических заболеваний (Государственный научно-исследовательский центр профилактической медицины Минздравсоцразвития РФ)
- Система оценки и мониторинга производства изделий микроэлектроники (Cadence Design Systems, Inc.)

Подробное описание технологии и разработанных систем доступно по следующей ссылке:

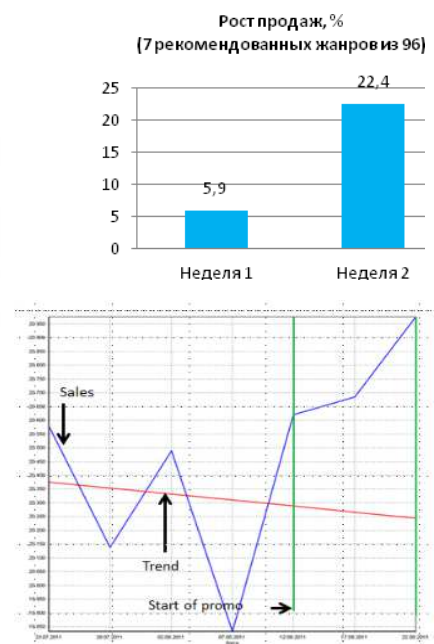
<http://intsys.msu.ru/staff/ryzhov/Systems%20for%20evaluation%20and%20monitoring%20of%20complex%20processes.pdf>

Разработанное решение имеет широкий спектр приложений и может использоваться в организациях от международного и федерального уровней до предприятий мелкого и среднего бизнеса.

2. Big Data/ Data Mining в телекоммуникационном бизнесе. Бизнес мобильного контента показывает впечатляющую динамику роста на протяжении многих лет. Использование методов интеллектуального анализа данных позволяет значительно повысить его эффективность. Обработка миллионов транзакций и выявление скрытых в них закономерностей поведения клиентов позволяет разрабатывать эффективные рекомендательные системы. Ниже представлены архитектура такого основанного на Big Data/ Data Mining рекомендательного инструмента и результаты его работы.



Разработанное решение может использоваться в розничной и интернет торговле.



3. Big Data/ Data Mining в торговле.

3.1. Профили клиентов.

Вход: база данных чеков + база данных дисконтной программы.

Вопрос: как устроен сегмент высокодоходных клиентов?

Решение: доходность (PROFIT)

разбита на 4 категории (меньше

8200; 8200 – 23300; 23300 – 60500; больше 60500). Строим профиль «PROFIT больше 60500».

Результат: люди, которые часто посещают магазин (в 3,5 раза чаще, чем в среднем по базе), возраста до 30 лет (в 3 раза чаще), покупают мало товаров, средняя сумма чека – маленькая, ...

	A	B	C	D	E	F	G
4	Если PROFIT is больше 60500 (max 351100)	то nBUY is больше 100 (max 3800)	25%	87%	3.5	1	
6	Если PROFIT is больше 60500 (max 351100)	то AGE is до 30	25%	38%	3.04	1	
19	Если PROFIT is больше 60500 (max 351100)	то SUM is меньше 100	25%	68%	2.58	1	
23	Если PROFIT is больше 60500 (max 351100)	то nGOODS is [1 - 2]	25%	63%	2.32	1	
40	Если PROFIT is больше 60500 (max 351100)	то SEX is 2	25%	75%	1.6	1	
104	Если PROFIT is больше 60500 (max 351100)	то TIME is [7-13]	25%	45%	1.25	1	
125	Если PROFIT is больше 60500 (max 351100)	то TYPECARD is 10%	25%	100%	1.2	1	
223	Если PROFIT is больше 60500 (max 351100)	то DAY is конец месяца	25%	39%	1.12	1	
224	Если PROFIT is больше 60500 (max 351100)	то MONTH is апрель	25%	29%	1.12	1	
307	Если PROFIT is больше 60500 (max 351100)	то DAY is пн.-чт.	25%	61%	1.08	1	
308	Если PROFIT is больше 60500 (max 351100)	то MONTH is весна	25%	57%	1.08	1	
309	Если PROFIT is больше 60500 (max 351100)	то MONTH is февраль	25%	26%	1.08	1	
369	Если PROFIT is больше 60500 (max 351100)	то MONTH is март	25%	28%	1.05	1	
512	Если PROFIT is больше 60500 (max 351100)	то DAY is середина месяца	25%	34%	1.02	1	
864							
865							

3.2. Поведение клиентов.

Вход: база данных чеков.

Вопрос: какие товары наиболее хорошо продаются в определенное время (года, недели, дня)?

Решение: разбиваем товары на

группы, разбиваем время на

периоды, применяем data mining.

Результат:

- для клиентов со средней суммой чека весной характерно приобретать товар группы №10 (в 2 раза чаще случайного поведения);

- для них же зимой характерно совместное приобретение товаров групп № 5, 7, 9;

- ...

	A	B	C	D	E	F	G	H
2661	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is зима	13%	2.11	47%	2	3	1
2662	Если SUM is [1000 - 40000] and MONTH is зима	то SUM is [1000 - 40000] and MONTH is весна	11%	2.11	63%	2	3	1
2663	Если MONTH is весна and nGOODS is [14 - 105]	то SEX is 1 and nGOODS is [14 - 105]	13%	2.11	54%	2	2	3
2664	Если SEX is 1 and nGOODS is [14 - 105]	то SEX is 2 and SUM is [1000 - 40000]	11%	2.1	27%	2	2	2
2665	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.1	35%	2	2	2
2666	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is зима	13%	2.1	31%	2	2	2
2667	Если SUM is [1000 - 40000] and TYPECARD is 10%	то SUM is [1000 - 40000] and TYPECARD is 10%	23%	2.1	47%	2	1	3
2668	Если SUM is [1000 - 40000] and TYPECARD is 10%	то SUM is [1000 - 40000] and TYPECARD is 10%	23%	2.1	47%	2	3	1
2669	Если SUM is [1000 - 40000] and TYPECARD is 10%	то SUM is [1000 - 40000] and TYPECARD is 10%	23%	2.1	63%	2	3	3
2670	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.09	42%	2	3	3
2671	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.09	30%	2	2	2
2672	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.09	35%	2	2	2
2673	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is весна	13%	2.09	63%	2	1	3
2674	Если SUM is [1000 - 40000] and MONTH is зима	то SUM is [1000 - 40000] and MONTH is зима	11%	2.09	47%	2	3	3
2675	Если MONTH is зима and nGOODS is [14 - 105]	то SEX is 2 and SUM is [1000 - 40000]	11%	2.09	60%	2	1	3
2676	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.08	63%	2	2	3
2677	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is весна	13%	2.08	48%	2	3	2
2678	Если SUM is [1000 - 40000] and MONTH is зима	то SUM is [1000 - 40000] and MONTH is зима	11%	2.08	47%	2	1	1
2679	Если DAY is пн.-чт. and nGOODS is [14 - 105]	то SEX is 2 and SUM is [1000 - 40000]	13%	2.08	47%	2	2	1
2680	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.07	42%	2	2	2
2681	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is весна	13%	2.07	35%	2	2	2
2682	Если SUM is [1000 - 40000] and MONTH is весна	то SUM is [1000 - 40000] and MONTH is весна	13%	2.07	46%	2	2	3
2683	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.06	46%	2	3	3
2684	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.06	46%	2	1	1
2685	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.06	31%	2	2	2
2686	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.06	30%	2	2	2
2687	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.06	40%	2	3	3
2688	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.05	26%	2	3	3
2689	Если SEX is 2 and SUM is [1000 - 40000]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.04	46%	2	1	1
2690	Если SUM is [1000 - 40000] and DAY is пн.-чт.	то SUM is [1000 - 40000] and DAY is пн.-чт.	12%	2.04	46%	2	1	1
2691	Если MONTH is зима and nGOODS is [14 - 105]	то SEX is 1 and nGOODS is [14 - 105]	11%	2.04	52%	2	1	1
2692	Если MONTH is зима and nGOODS is [14 - 105]	то SEX is 2 and SUM is [1000 - 40000]	11%	2.04	46%	2	1	1

2.3. Совместные покупки.

Вход: база данных чеков

Вопрос: какие товары хорошо продаются вместе?

Решение: разбиваем товары на

группы, применяем data mining

Результат:

- группа #14 хорошо продается с товарами групп №5, 4, 7, 9 (почти в 4 раза чаще по сравнению с независимыми покупками);
- добавление товаров групп № 6 или 13 увеличивает продажи;
- ...

Решения могут быть использованы в торговых организациях от крупных сетей до малого бизнеса.

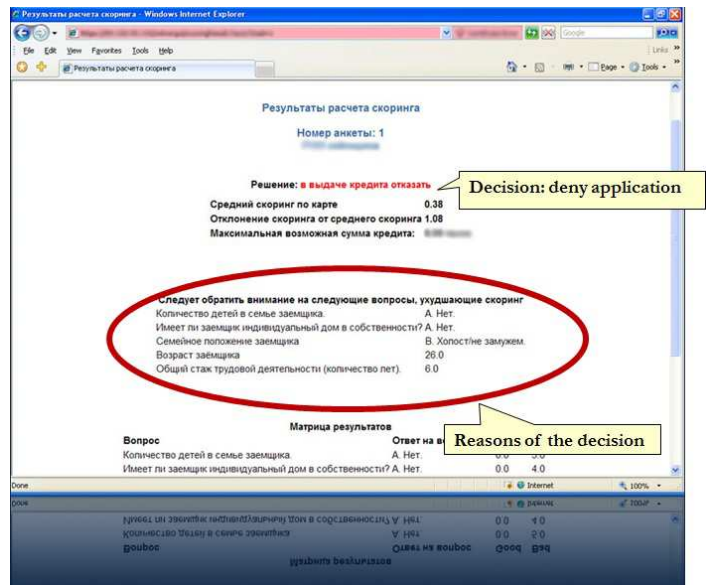
	A	B	C	D	E	F	G	H
1	Если	то	Поддержка	Интерес	Доля	Длина	Длина следствия	
2	Если GOODS is [14]	то GOODS is [14]	24%	3.64	26%	1	5	
3	Если GOODS is [14]	то GOODS is [14]	24%	3.62	26%	1	5	
4	Если GOODS is [14]	то GOODS is [14]	24%	3.59	28%	1	5	
5	Если GOODS is [14]	то GOODS is [14]	24%	3.59	27%	1	5	
6	Если GOODS is [14]	то GOODS is [14]	24%	3.58	28%	1	5	
7	Если GOODS is [14]	то GOODS is [14]	24%	3.57	26%	1	5	
8	Если GOODS is [14]	то GOODS is [14]	24%	3.56	28%	1	5	
9	Если GOODS is [14]	то GOODS is [14]	24%	3.51	27%	1	4	
10	Если GOODS is [14]	то GOODS is [14]	24%	3.49	28%	1	4	
11	Если GOODS is [14]	то GOODS is [14]	24%	3.49	28%	1	4	
12	Если GOODS is [14]	то GOODS is [14]	24%	3.49	28%	1	4	
13	Если GOODS is [14]	то GOODS is [14]	24%	3.49	28%	1	4	
14	Если GOODS is [14]	то GOODS is [14]	24%	3.48	31%	1	5	
15	Если GOODS is [14]	то GOODS is [14]	24%	3.47	25%	1	4	
16	Если GOODS is [14]	то GOODS is [14]	24%	3.47	33%	1	5	
17	Если GOODS is [14]	то GOODS is [14]	24%	3.46	30%	1	4	
18	Если GOODS is [14]	то GOODS is [14]	24%	3.46	29%	1	4	
19	Если GOODS is [14]	то GOODS is [14]	24%	3.45	29%	1	4	
20	Если GOODS is [14]	то GOODS is [14]	24%	3.45	30%	1	4	
21	Если GOODS is [14]	то GOODS is [14]	24%	3.45	29%	1	4	
22	Если GOODS is [14]	то GOODS is [14]	24%	3.44	29%	1	4	
23	Если GOODS is [14]	то GOODS is [14]	24%	3.44	27%	1	4	
24	Если GOODS is [14]	то GOODS is [14]	24%	3.44	28%	1	4	
25	Если GOODS is [14]	то GOODS is [14]	24%	3.43	29%	1	4	
26	Если GOODS is [14]	то GOODS is [14]	24%	3.43	35%	1	5	
27	Если GOODS is [14]	то GOODS is [14]	24%	3.42	29%	1	4	
28	Если GOODS is [14]	то GOODS is [14]	24%	3.42	30%	1	4	
29	Если GOODS is [14]	то GOODS is [14]	24%	3.42	27%	1	4	
30	Если GOODS is [14]	то GOODS is [14]	24%	3.42	28%	1	4	
31	Если GOODS is [14]	то GOODS is [14]	24%	3.39	26%	1	3	
32	Если GOODS is [14]	то GOODS is [14]	24%	3.39	34%	1	4	
33	Если GOODS is [14]	то GOODS is [14]	24%	3.38	25%	1	3	
34	Если GOODS is [14]	то GOODS is [14]	24%	3.38	33%	1	4	
35	Если GOODS is [14]	то GOODS is [14]	24%	3.38	26%	1	4	
36	Если GOODS is [14]	то GOODS is [14]	24%	3.36	31%	1	4	

4. Big Data/ Data Mining в банковском бизнесе: системы кредитного скоринга.

Вход: анкеты + кредитные истории заемщиков. Методика оценки качества кредитов. Технология Data mining.

Выход: интегрированная система автоматической оценки заемщиков.

Результат: в 2 раза меньше дефолтов, в 5 раз меньше «плохих» кредитов.



5. Big Data/ Data Mining в финансах: выявление подозрительных транзакций.

Вход: база транзакций.

Технология: искусственные нейронные сети, нечеткая логика, кластеризация.

Решение:

- Автоматическое выявление нетипичных транзакций.
- Автоматическое выявление транзакций, похожих на указанные пользователем.

Результат: В 7 раз больше «плохих» транзакций по сравнению со случайной выборкой.

Дата	Описание	Операция	Счет	Сумма	Счет	Сумма	С/Другое	Сальдо
30.09.04	Бухгалтерия	Списание на счет	722	91.1	320.83	€		136.714.00
21.09.04	Вилкина 000502	Денежные на руб.	91	91.1	472.50	€		10.884.468.42
30.09.04	Бухгалтерия	Списание на счет	722	91.1	300.00	€		9.295.163.90
06.11.04	Вилкина 000549	Денежные на руб.	91	91.1	412.87	€		11.388.468.37

6. Big Data/ Data Mining в HR:

6.1. Анализ резюме, оценка персонала.

Вход: текстовая база резюме.

Вопросы: Кто лоялен, кто нет? Кто растет, а кто стоит на месте? Как собрать эффективную проектную команду? Быстрый отбор кандидатов на должность.

Технология: Ассоциативные правила, нечеткая логика.

Выход: профили лояльных/ эффективных/ и других типов сотрудников, индикаторы лояльности, эффективности, и других качеств (в зависимости от компании).

Если	То	Интерпретация	Доля	Длина
если ср. время работы есть меньше 1	то => ин. язык есть French		2.73	30%
если ср. время работы есть меньше 1	то => пол есть женский		2.62	33%
если ср. время работы есть меньше 1	то => отдел есть marketing		2.11	11%
если ср. время работы есть меньше 1	то => отдел есть sales		1.72	11%
если ср. время работы есть меньше 1	то => отдел есть clients		1.69	26%
если ср. время работы есть меньше 1	то => ин. язык есть German		1.61	22%
если ср. время работы есть меньше 1	то => ин. язык есть english		1.55%	1
если ср. время работы есть меньше 1	то => кол-во компаний			
если ср. время работы есть меньше 1	то => кол-во компаний			
если ср. время работы есть меньше 1	то => университет есть			
если ср. время работы есть меньше 1	то => кол-во компаний			
если ср. время работы есть меньше 1	то => факультет есть			

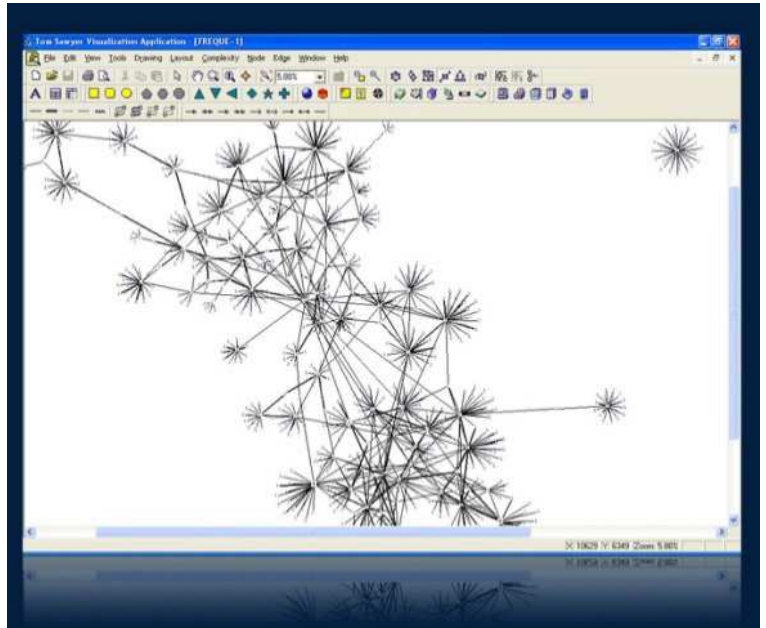
French-speaking women from sales or marketing department.

6.2. Анализ и оптимизация коммуникационной структуры компании

Вход: логи корпоративной АТС, exchange server, других используемых ИТ инструментов.

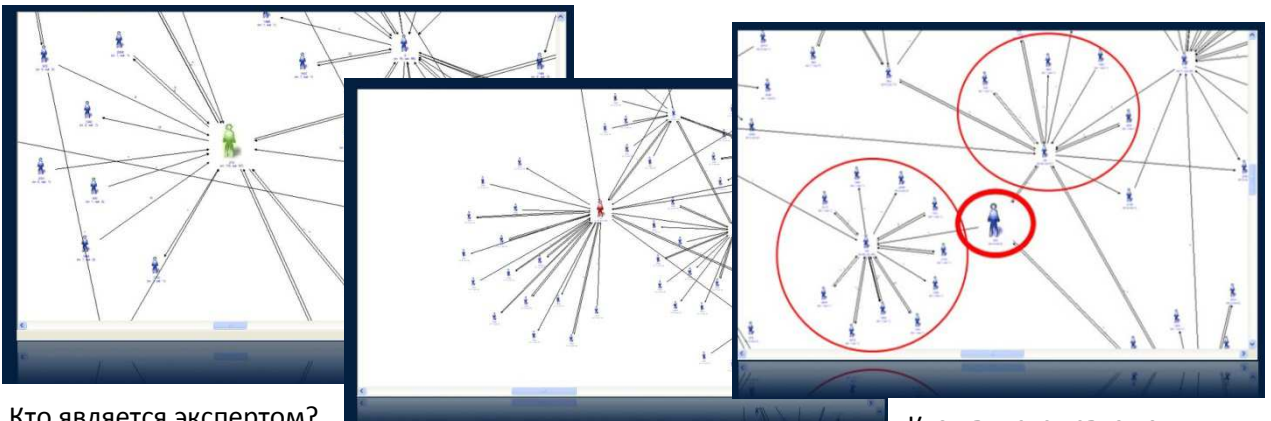
Вопросы: Выявление в компании неформальных группировок и их лидеров:

- Кто является экспертом? К кому обращаются люди?
- Кто является инициатором?
- Кто является связующим звеном между группами?
- Как повысить эффективность коммуникаций и распространение знаний внутри компании/ подразделения/ проектной группы?



Технология: Кластеризация, нечеткая логика, анализ социальных сетей.

Выход: рекомендации по повышению эффективности коммуникаций и распространению знаний



Кто является экспертом?

Как распространяется информация/ знания?

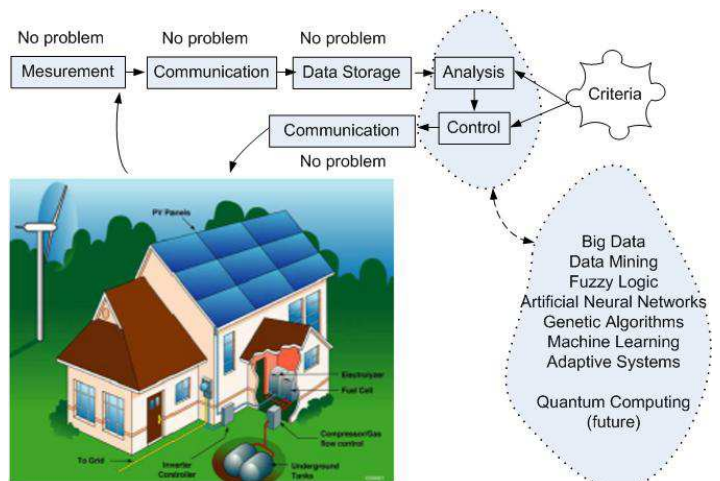
Кто является связующим звеном между группами?

Разработанные решения могут использоваться для оптимизации и повышения эффективности работы организаций крупного, среднего и мелкого бизнеса.

7. Big Data/ Data Mining в энергетике: интеллектуальные сети.

Вход: показания потребления энергии домашними приборами/ домовладениями в рамках подстанции.

Вопросы: Как управлять умным счетчиком/ подстанцией для обеспечения минимизации скачков напряжения (повышения качества электричества)?

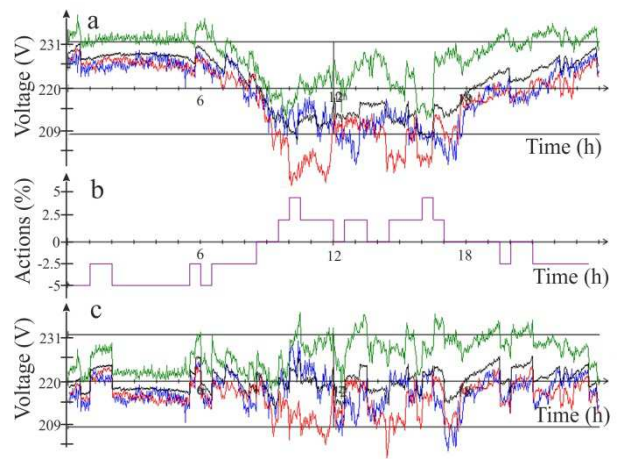


Технология: Ассоциативные правила, нечеткая логика, машинное обучение.

Выход: адаптивный алгоритм управления умным счетчиком/ подстанцией.

Результат: на реальных данных повышение качества электричества в десятки раз.

Решение может быть использовано для оптимизации энергосетей от домохозяйств и дачных участков до городского и регионального уровней.



Суммируя опыт, полученный в ходе выполнения десятков проектов в организациях различного масштаба (от международного и федерального уровней до саппапов) и отраслевой специализации в течение более 25 лет, можно утверждать, что методы интеллектуально анализа данных (Big Data в современной терминологии) являются действенным инструментом повышения эффективности любого бизнеса.

Общая схема использования инструментов Big Data/ Data Mining в бизнесе:

Бизнес – это клиенты

На ком и как мы зарабатываем? На ком и как теряем?

Инструменты Big Data/ Data Mining позволяют обрабатывать транзакции и получать ответы на бизнес-вопросы.

ИТ-инфраструктура позволяет накапливать транзакции.

