

*Сидоренко А.А.,
студентка магистратуры,
2 курс, Институт экономики, математики
и информационных технологий,
факультет информационных технологий и анализа данных,
отделение Бизнес-информатики,
Российская академия народного хозяйства и государственной службы
при Президенте Российской Федерации,
Россия, г. Москва*

*Научный руководитель: Уманский Ю.Н., кандидат военных наук, доцент,
доцент Кафедры системного анализа и информатики,
Российская академия народного хозяйства и государственной службы
при Президенте Российской Федерации,
Россия, г. Москва*

**СНИЖЕНИЕ ПОТЕРЬ ПРИ ПРОСТОЕ ЭНЕРГОБЛОКОВ АЭС ЗА СЧЕТ
ПРИМЕНЕНИЯ DATA MINING И ПРЕДИКТИВНОЙ АНАЛИТИКИ ПРИ
ПРОГНОЗИРОВАНИИ АВАРИЙНЫХ СИТУАЦИЙ В РАБОТЕ
ЭНЕРГОБЛОКА**

Аннотация: В статье рассматривается вопрос целесообразности применения методов Data mining и предиктивной аналитики при прогнозировании аварийных ситуаций в работе АЭС. Описываются особенности структурированных и неструктурированных данных на производстве, сложность и различия в их обработке. Особый акцент сделан на важность неструктурированных данных на производстве и их роль при прогнозировании событий.

Ключевые слова: интеллектуальный анализ данных, предиктивная аналитика, неструктурированные данные, прогнозирование аварийных событий на производстве.

Annotation: *The article addresses the issue of the feasibility of applying Data mining methods and predictive analytics in forecasting emergency situations during NPP operation. It describes the features of structured and unstructured data in production process, the complexity and differences in their processing. Particular emphasis is placed on the importance of unstructured data in production process and their role in forecasting events.*

Key words: *Data mining, predictive analytics, unstructured data, forecasting of emergency events in production.*

На сегодняшний день АО «Концерн Росэнергоатом» - единственная компания, эксплуатирующая все российские атомные электростанции. Данная деятельность сопряжена с огромной ответственностью перед обществом и государством. Основными приоритетами компании являются: энергетическая безопасность, защищенность и безопасность граждан и окружающей среды, а также экономическое развитие России. Для снижения потерь при простое энергоблока АЭС, связанного с аварийной остановкой и выводом в ремонт, необходим постоянный учет и анализ опыта эксплуатации станций, расследование причин аномальных и аварийных событий в работе. Разработка и реализация превентивных, корректирующих и компенсирующих мероприятий на основании анализа внутреннего и внешнего опыта эксплуатации, а также контроль выполнения и оценка результативности разработанных превентивных, корректирующих и компенсирующих мероприятий.

Однако данный процесс осложняется тем, что порядка 70% всех данных, связанных с работой АЭС – это неструктурированные данные, т.е. не имеющие заранее определенную структуру, либо не организованные в установленном порядке. Часто подобные данные представлены в формате электронного документа, содержащего даты, различные факты и показатели. В таких

ситуациях применение традиционных программ, предназначенных для работы со структурированными данными (хранящимися в базах или аннотированными) не подходит. В целом затрудняется анализ информации, увеличивается время подготовки экспертных заключений и время принятия управленческих решений, усложняется процесс планирования превентивных мероприятий и в целом увеличиваются трудозатраты по процессу.

Сам по себе вопрос анализа неструктурированных данных не является новым. Однако в эпоху «Big Data» данная тема все чаще поднимается в производственных компаниях, имеющих огромный массив данных с информацией по накопленному опыту. Темпы роста хранимой информации и разнообразие обрабатываемой информации продолжают значительно расти, усложняя процесс управления данными.

Структурированные данные в отличие от неструктурированных представляют собой совокупность предсказуемых, повторяющихся и упорядоченных единиц информации. Такие данные можно разложить в строки и столбцы, размечать и проводить по ним анализ, но таких данных в компании всего порядка 30-35%. Неструктурированные данные в свою очередь нельзя разложить в виде классической базы данных структурированного типа. Как правило они хранятся в виде электронных писем, отчетов, актов и протоколов расследований аварийных ситуаций на производстве, предписаний уполномоченных органов, а также в виде данных, генерируемых IoT устройствами. Обобщенно можно сказать, что работа со структурированными данными дает ответ на вопрос «что происходит?», а с неструктурированными «почему это происходит?».

Производственные компании, игнорирующие в своей работе накопленную неструктурированную информацию, теряют существенное количество данных, остающихся вне поля аналитических инструментов, которое потенциально могло бы предотвратить аварийные ситуации в основной деятельности и подтолкнуло бы организацию к росту. Простой пример: показатели состояния защит энергоблока, показатели активной зоны - это структурированные данные,

а неструктурированные несут более глубокую смысловую нагрузку, позволяют понять в каком режиме эксплуатируется энергоблок, проанализировать какие события происходят при такой нагрузке и почему.

По прогнозам крупных исследовательских компаний, к 2022 году общий процент неструктурированных цифровых данных составит 92% от общего числа данных, что серьезно повлияет на процессы управления данными и на подходы к их защите. Почти все производственные компании ведут учет аварийных и нештатных ситуаций, с их подробным описанием, расследованием причин и принятыми мероприятиями, но мало кто способен оперативно проанализировать всю историю таких событий при возникновении рисков их наступления и разработать план действий. Знать почему данное событие происходит бывает недостаточно, важно понимать может ли оно повториться. Именно в таких ситуациях и применяются методы Data mining.

Data mining – совокупность методов по «добычи» неизвестных ранее, нестандартных, полезных, с практической точки зрения, знаний. Данные методы могут применяться и к неструктурированным данным, их применение обеспечит анализ неструктурированной текстовой или графической информации, а также формирование различных прогнозов. Полученные знания можно использовать для предотвращения аварийных ситуаций в работе АЭС, обоснования проведения внеплановых ремонтов, своевременного внесения корректировок в плановые ремонты, внесения изменений в планы закупок материалов для ремонта и совершенствования производственных процессов в целом.

В конструкции энергоблоков с реактором типа РБМК (реактор большой мощности канальный) предусмотрено два турбогенератора. Данная технология позволяет вывести один турбогенератор в ремонт без останова самого реактора, тем самым обеспечивая бесперебойную работу реактора и выработку электроэнергии. В случае наличия прогноза аварийной ситуации, а существующие математические методы позволяют строить прогнозы за 2-4 месяца до ее возникновения, станция сможет подготовиться и провести

внеплановый ремонт без потери выручки и нарушения плана ФАС (Федеральной антимонопольной службы).

Прогнозные модели строятся с применением методов углубленного анализа данных и различных математических алгоритмов: искусственные нейронные сети, регрессионный анализ, корреляционный анализ, деревья решений и др. Эти методы также заложены в основу информационно-аналитических систем, которые объединяют, хранят и анализируют информацию из различных источников.

Для того чтобы делать прогнозы, необходимо на первом этапе осуществить поиск и анализ большого массива неструктурированных данных. Эти задачи должны быть реализованы в виде подсистемы, которая также будет:

- осуществлять интеллектуальный поиск по содержанию документов, с учетом морфологии русского языка, по ключевым словам;
- осуществлять параметрический поиск (поиск и выдача значений отдельных полей документов);
- обеспечивать возможности поиска ключевых слов внутри фильтров (категорий) для быстрого доступа к конкретным фильтрам, автоматическое формирование фильтров поиска в зависимости от содержимого документов;
- обеспечивать возможности поиска схожих документов;
- осуществлять автоматическое формирование перечня схожих по содержимому документов;
- формировать структуры индексирования документов по отдельным разделам, метаполям документа;
- фильтровать результаты поиска на основе прав доступа пользователей, осуществляющих поиск (разграничение доступа к информации на уровне доступа к вкладкам интерфейса пользователя и логических баз данных);
- обеспечивать возможность индексации новых документов и данных без необходимости приостанавливать функции поиска информации;

- осуществлять нечеткий поиск (для неточно заданных грамматически ключевых слов);
- осуществлять выявление ошибок в написании слов в рамках поисковых запросов (в строке поиска) и автоматическое формирование предложений по корректировке запросов;
- осуществлять автоматическое формирование ключевых понятий документа;
- формировать механизмы создания справочников сущностей (ключевых слов) с автоматическим их выделением в обрабатываемых документах на основе фиксированного количества справочников и их согласованного внутреннего состава;
- осуществлять распознавание графических документов: решать задачи по обработке изображений, на основе оптического распознавание символов.

Одной из особенностей АО «Концерн Росэнергоатом» является территориальная распределённость АЭС, Центрального аппарата и контролирующих государственных органов, что дополнительно усложняет сбор и обработку структурированных и неструктурированных данных, важных для процесса анализа опыта эксплуатации. В то время, как накопленный опыт одной станции, может быть критически важен для принятия решений на другой АЭС.

Дополнительны факторы, затрудняющие процесс анализа:

- существенные различия в программном построении информационных баз данных (информационных ресурсов) и процессов сбора информации по опыту эксплуатации на разных АЭС;
- отсутствие единой системы регистрации событий эксплуатации АЭС;
- отсутствие единой точки доступа к данным об опыте эксплуатации АЭС, событиях, замечаниях и предложениях в условиях разрозненности информационных систем;

- отсутствие единой информационной системы управления процессом расследований и документированием результатов;
- сложность обмена информацией между функциональными группами специалистов географически распределенных филиалов;
- отсутствие единого источника данных о надежности оборудования и эффективности стратегий ремонтов с учетом конструктивных особенностей блоков и конфигурации оборудования;
- отсутствие единого источника данных по поставщикам оборудования и поставщикам услуг по проведенным ремонтам энергоблоков.

Исходя из этого в основу системы должно быть заложено единое логическое пространство Больших данных, т.е. выстроена единая платформа. Разработка единой платформы, включающей центр обработки и хранения данных, позволит унифицировать интерфейсы взаимодействия между предприятиями АО «Концерн Росэнергоатом» и внешней средой, тем самым реализовав сквозные бизнес-процессы энергетического дивизиона.

В компании также есть система управления производством MES-класса (Manufacturing Executions System), предоставляющая актуальные данные реального времени обо всех аспектах работы станций. Аналитика в режиме реального времени по неструктурированным данным и данным реального времени - это база для предиктивного анализа и принятия решений, повышения производительности труда, сокращения времени протекания процессов и снижения себестоимости продукции.

Важным аспектом методов «добычи» полезных знаний, с практической точки зрения, является представление результатов, в наглядном и понятном для конечного пользователя виде, в терминах предметной области, не требующих дополнительного разъяснения. Для этого создается подсистема, способная:

- формировать различные визуальные формы представления данных;
- предоставлять данные в наглядном, интуитивно понятном виде (различные шкалы, показатели, индикаторы и т. п.);

- выгружать отчеты по предиктивной аналитике в стандартных форматах;
- предоставлять возможность преднастройки отчетов пользователям (в том числе возможность конструирования отчетных таблиц по различным критериям анализа);
- осуществлять цветовое кодирование отображения показателей в зависимости от установленных предельных значений;
- отображать исторические значения показателей;
- и т.д.

Ниже представлены два варианта визуального отображения графиков преднастроенного отчета. (Рисунок 1, Рисунок 2, Рисунок 3)

Способы визуализации графиков преднастроенного отчета с одновременным отображением детализации по одной аналитике:

- динамика показателя, объект типа «Линейная диаграмма», изменение значения во времени;
- тренд показателя, объект типа «Линейная диаграмма», накопительный итог показателя во времени;

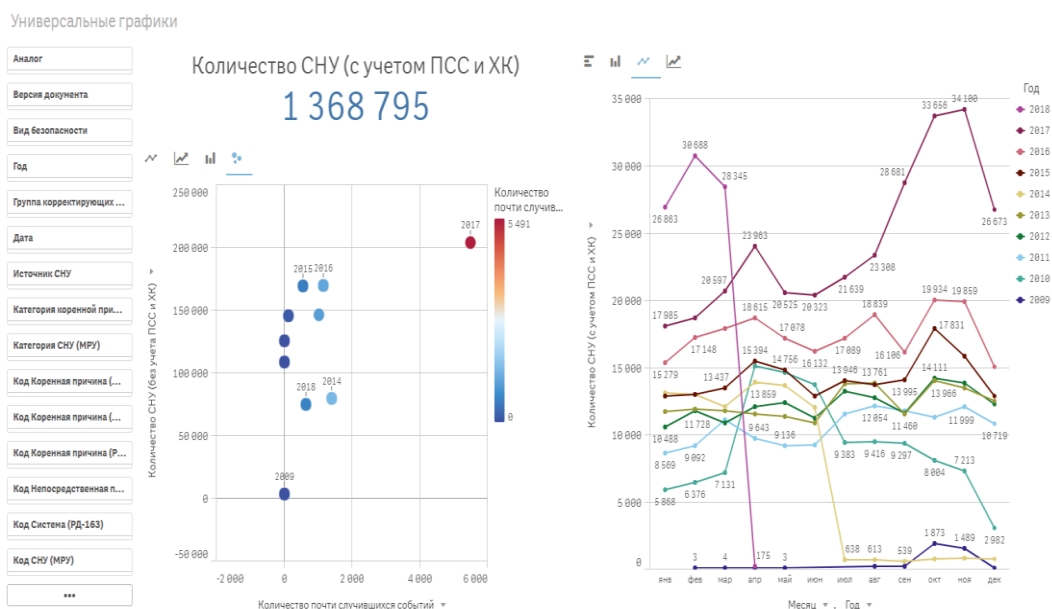


Рисунок 1 Пример преднастроенного отчета (Линейная диаграмма)

Конструктор таблиц позволяет построить таблицу по всем показателям и критериям анализа, собранными системой, разложив все данные по максимально возможной детализации построчно. (Рисунок 4)

The screenshot shows a web application interface titled 'Конструктор: Табличный вид'. It features a table with columns: 'Квартал', 'Название корректирующего мероприятия', 'Доля задач', and 'Количество задач'. The table contains two rows of data for 'КВ1'. A sidebar on the left lists various filters and measures, with 'Квартал', 'Доля задач', and 'Количество задач' highlighted in green. The table header includes filters for 'Квартал', 'Название ко...', 'Доля задач', and 'Количество з...'.

Квартал	Название корректирующего мероприятия	Доля задач	Количество задач
КВ1	1. При вводе в работу нового оборудования (кроме УРЗА) проводить инструментальный контроль (тепловизионный, ультразвуковой и др.) в течение 24 часов после подачи напряжения и через 6-12 часов после	0,1%	10
КВ1	1. Провести анализ схем гальванической развязки входов блоков приемыки внешних дискретных сигналов (команд) микропроцессорных регуляторов возбуждения при наличии длинных кабельных линий. При	0,1%	9

Рисунок 4 Конструктор таблиц

Вариативное представление результатов анализа позволяет расширить круг пользователей системы и подготовить отчеты различного назначения. В целом система повышает оперативность доступа к материалам по значимым событиям при расследовании аномальных событий (нарушения, отклонения).

Подводя итог, можно сказать, что управление производством и принятие решений в современных условиях затруднительны, а порой невозможны, без качественной и оперативной информации о текущем состоянии производственных активов и прогнозов аварийных событий.

Подсистема поиска и анализа неструктурированных данных, с заложенными в нее методами Data Mining, Text Mining и предиктивного анализа, снижает временные и финансовые издержки на поиск и анализ значимых событий опыта эксплуатации, а также повышает качество анализа данных опыта эксплуатации на производстве, позволяет получить достоверные прогнозы, повышает оперативность и эффективность принимаемых управленческих

решений. Ниже приведен график недовыработки электроэнергии, связанный с отказами генераторов до применения методов Data Mining в прогнозировании аварий и после. (Рисунок 5).

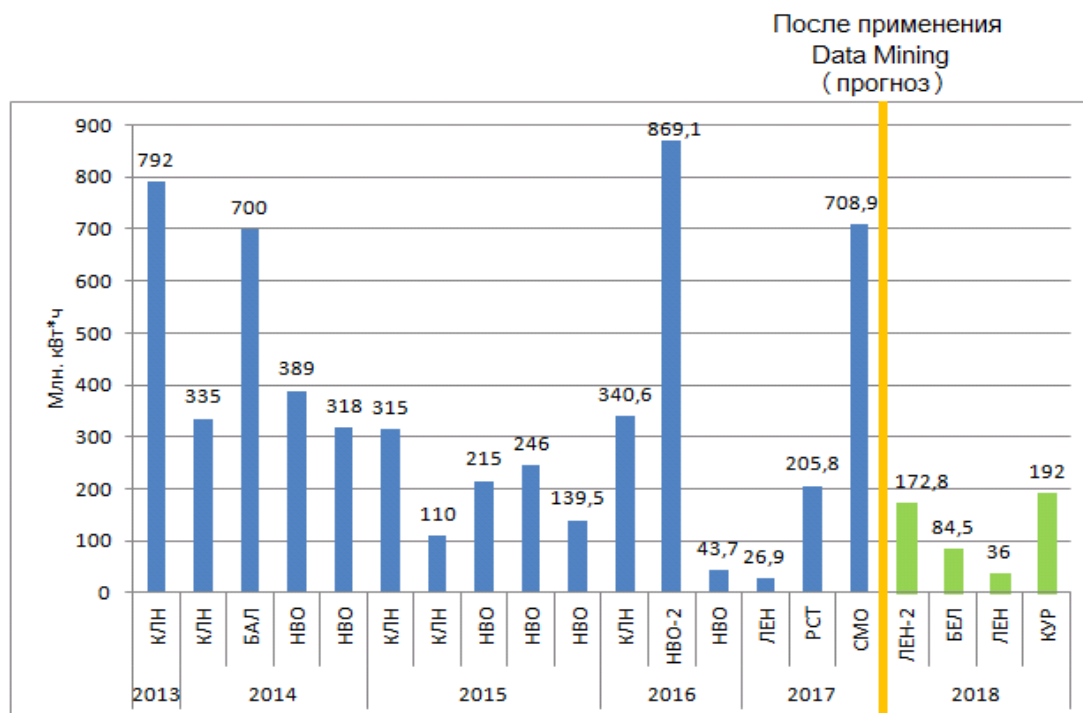


Рисунок 5 Недовыработка электроэнергии, млн. кВт*ч

Инструменты визуализации полученных результатов анализа помогут повысить эффективность и достоверность информационной поддержки персонала Концерна при принятии своевременных управленческих решений, нацеленных на обеспечение безопасности, надежности и эффективности деятельности.

Консолидация информации в одном месте о работе всей компании позволит расширить варианты для принятия оперативных решений и снизит в разы время их принятия, в конечном счете повысит безопасность на производстве.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Пятецкий-Шапиро Г., Data Mining и перегрузка информацией // Вступительная статья к книге: Анализ данных и процессов / Барсемян А. А.,

Куприянов М. С., Холод И. И., Тесс М. Д., Елизаров С. И.. 3-е изд. перераб. и доп. СПб.: БХВ-Петербург, 2009. 512 с.

2. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям (+ CD). — СПб.: Изд. Питер, 2009. — 624 с.

3. Журавлёв Ю.И., Рязанов В.В., Сенько О.В. РАСПОЗНАВАНИЕ. Математические методы. Программная система. Практические применения. — М.: Изд. «Фазис», 2006. — 176 с. — ISBN 5-7036-0108-8.

4. Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. — 3rd Edition. — Morgan Kaufmann, 2011. — P. 664. — ISBN 9780123748560.

5. Пескова О. В. Алгоритмы классификации полнотекстовых документов // Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. — М.: МИЭМ (Московский государственный институт электроники и математики), 2011. — С. 170—212. — ISBN 978-5-94506-294-8.

6. Survey of Text Mining I: Clustering, Classification, and Retrieval / Ed. by Berry M. W. — 2004. — Springer, 2003. — 261 p. — ISBN 0387955631.

7. Aggarwal C. C., Zhai C. Mining Text Data. — Springer, 2012. — 527 p. — ISBN 9781461432234.

8. Do Prado H. A. Emerging Technologies of Text Mining: Techniques and Applications / Ed. by H. A. Do Prado, E. Ferneda. — Idea Group Reference, 2007. — 358 p. — ISBN 1599043734.