

# Применение прикладного программного пакета STATISTICA (StatSoft) для решения практических задач прогнозирования в области энергетики

**Технический директор StatSoft Russia**  
**Милков Максим**

[expert@statsoft.ru](mailto:expert@statsoft.ru)

04/12/2014

# О компании

# О компании StatSoft

**Многолетний опыт  
в аналитике  
с 1984 года**



**Продукты StatSoft  
переведены на 12  
языков**



**Авторизованные  
представительства  
в 20 странах мира**



**StatSoft®**



**Продукты  
StatSoft  
включают в себя  
более 16 000  
аналитических  
процедур и  
методов**

**Более 1 000 000  
легальных  
пользователей**



**Обширная  
сфера  
применения  
продуктов  
STATISTICA**

# Наши клиенты



Центральный  
банк  
Российской  
Федерации

**StatSoft**®



**СеверСталь**  
открытое акционерное общество



**МТС**

**ИНТЕР**  **РАОЕЭС**

**ИНГОССТРАХ**  
*Ingosstrakh*



**ЛУКОЙЛ**  
НЕФТЯНАЯ КОМПАНИЯ



**РУСАЛ**

 **X5RETAILGROUP**



**phs** **Pharmstandard**



**НИЖФАРМ**®

# Деятельность StatSoft Russia



## Техническая поддержка

- Установка и оптимальная настройка системы
- Консультации по модулям программы
- Взаимодействие с другими приложениями
- Математические алгоритмы анализа



## Курсы лекций StatSoft

- Стандартная или индивидуальная программа
- Использование Ваших данных
- Выезд на Вашу территорию



## Консалтинг

- Постановка задачи, вербальное описание
- Методология и технология решения задачи
- Программная реализация
- Подготовка технической документации
- Обучение пользователей

# Реализованные проекты



**Система контроля качества**



**Система технической диагностики оборудования**



**Прогнозирование стоимости/затрат на строительство**



**Система управления техническим состоянием электровоза**



**Анализ и прогнозирование цен на электроэнергию**

**и множество других...**

# Данные

# Где данных много?

## Датчики

Высокотехнологичные изделия  
Мобильные устройства

1

## Системы учета

Биллинг  
ERP  
CRM

2

## Полу- и неструктурированные данные

Web-страницы  
Социальные сети

3

# У кого данных много?

Телеком



Ритейл



Финансовые  
организации



Поисковики  
Социальные  
сети



Производство



# Прогнозирование в энергетике

# Задача

Построение прогноза для всевозможных временных рядов:

**Цены**

**Объемы**

Особенности:

**Разные  
источники  
данных**

**Разная  
детализация**  
час-сутки-неделя...

**Различные  
горизонты**

# Проблемы?!

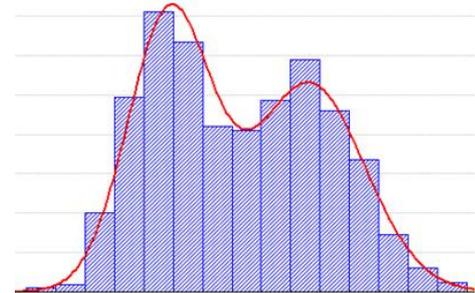
**Чистка  
данных**

**Подбор  
модели**

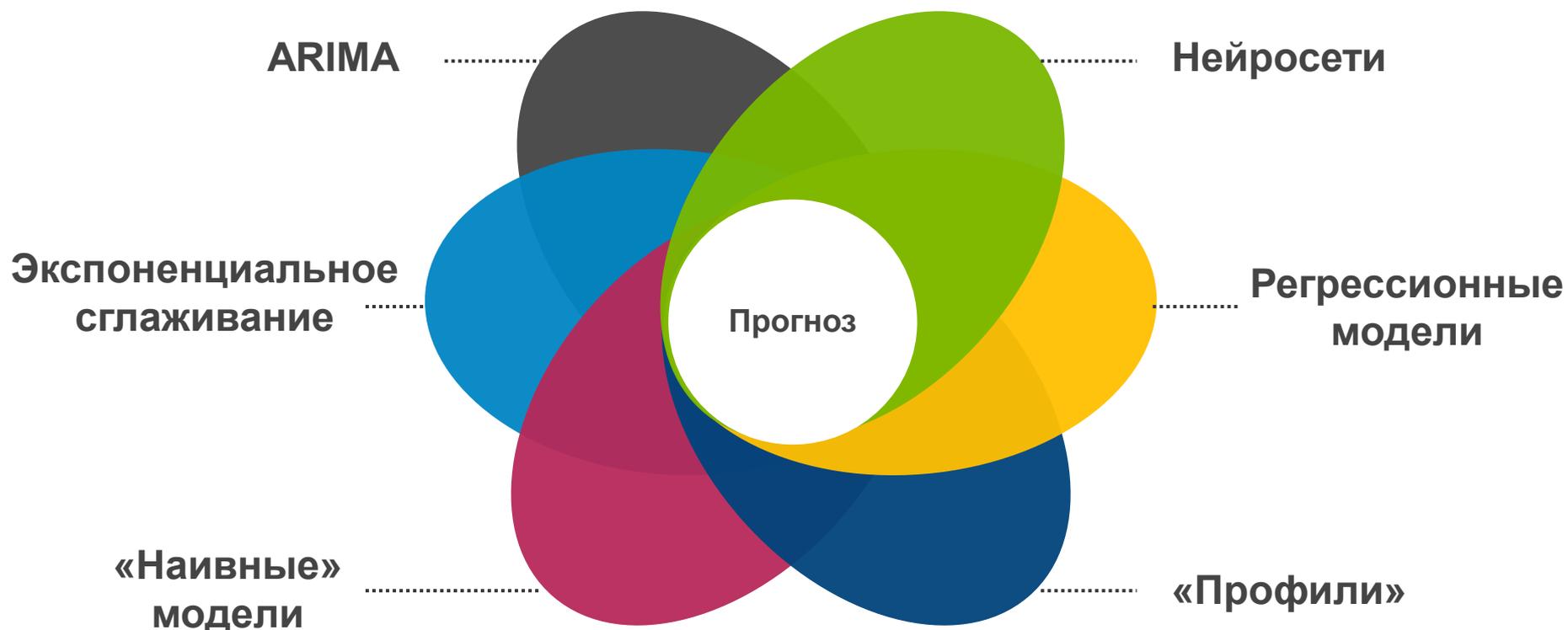
**Праздники**

# Чистка данных

- «ручное регулирование»
- **Нестационарность, мультимодальность данных за длительный период**



# Прогнозирующие модели



# Выбор модели

1.

- Отбор подходящих моделей на этапе разработки экспертом

2.

- Разбивка данных на обучающее и тестовое множество

3.

- Обучение моделей различного типа и различных «конфигураций»

4.

- Оценка точности модели на тестовом множестве (выбор метрики)

5.

- Выбор (ансамбля) модели-«победителя»

6.

- Переобучение «победителя(ей)» на всех данных

7.

- Построение прогноза

8.

- Оценка адекватности

# Праздники

## Календарь

- **Рабочие/нерабочие дни**

**или**

- **Регулярные тематические праздники**
  - **Новый год**
  - **8 марта и т.д.**
- **Нерегулярные праздники, события**
  - **Пасха**
  - **Игры сборной**

# Коррекция прогноза

1.

- В исторических данных на этапе чистки удаляются значения в период «нетипичных» дней

2.

- Образовавшиеся «окна» заполняются прогнозными значениями – получаем «рутинный» ряд

3.

- Для каждого типа события календаря строится модель зависимости Факт – «Рутинна»

4.

- Прогнозируется «рутинный» ряд

5.

- На основе моделей корректируются прогнозы в «нетипичные» дни

Что дальше

# Интеллектуальные счетчики

- **Больше рядов для прогнозирования**
- **Персонализированный анализ потребления**
  - **услуги по предоставлению информации третьим лицам (при получении согласия)**

# Учет факторов

- **Метеофакторы**
  - **Температура (зачастую недостаточно)**
  - **???**
- **Данные других отраслей**
  - **Телеком**
    - **например, количество активных абонентов в конкретном районе, в определенный интервал времени**

**Спасибо за внимание!**  
**Вопросы?**