

Санкт-Петербургский государственный университет
Математическое обеспечение и администрирование информационных систем
Информационно - аналитические системы

Волобой Валерия Сергеевна

Разработка и применение набора инструментов для прогнозирования продаж нефтепродуктов на АЗС

Бакалаврская работа

Научный руководитель:

к. ф.-м. н., доцент Михайлова Е. Г.

Рецензент:

начальник отдела систем анализа
прогнозирования,

ОАО «Газпром нефть»

Черницын И.Г.

Санкт-Петербург
2016

SAINT-PETERSBURG STATE UNIVERSITY
Software and Administration of Information Systems
Analytical Information Systems

Valeriya Voloboy

**Development and application of a tool set for forecasting
petroleum products sales at gas stations**

Bachelor's Thesis

Scientific adviser:

Associate Profesor E. G. Mikhalova

Reviewer:

Head of Department of system

Analysis and forecasting, Gazpromneft

I.G. Chernitsyn

Saint-Petersburg
2016

Оглавление

1. Введение.....	4
2. Терминология	6
3. Постановка задачи	7
4. Описание и подготовка данных	9
5. Обзор.....	10
5.1. Подходы управления данными	10
5.2. Методы прогнозирования	11
5.3. Существующий кейс	13
5.4. Связанные работы.....	14
6. Используемые алгоритмы	15
6.1. Holt-Winters.....	15
6.2. ARIMA	16
6.3. ETS	17
6.4. Нейронные сети	18
7. Ансамблирование и накопление статистики.....	18
8. Скользящий контроль.....	19
9. Эксперименты	20
9.1. Holt-Winters.....	20
9.2. ARIMA	21
9.3. ETS	22
9.4. Нейронные сети	23
10. Результаты.....	24
11. Заключение	25
12. Список литературы	27
13. Приложение	29

1. Введение

Грамотное прогнозирование продаж является одной из самых важных концепций успеха почти любой компании. Для того чтобы наиболее эффективно вести бизнес, необходимо правильно рассчитать прогноз, что позволит компании оптимизировать и контролировать расходы. Когда же речь идёт о продукции, то качественно проведённое прогнозирование объёмов продаж, позволит так же сформировать оптимальный запас, что не повлечёт за собой нехватки или недостачи продукта.

По данным прошлого года, вклад нефтепродуктов в ВВП России составлял около 30%, в доходы бюджета — около 50%, в объём российского экспорта — около двух третей. Существует мнение о том, что Россия «сидит на нефтяной игле». Сторонники этого мнения утверждают, что экономика России основывается почти целиком на экспорте углеводородов, и если объёмы этого экспорта снизятся или цена на нефть существенно упадёт, то это приведёт к полному экономическому краху России. Однако есть и противники этого мнения, поэтому не стоит всецело верить и думать, что в сегодняшней ситуации, когда цены на нефть весьма нестабильны, нашу страну ждёт неминуемый провал.

С другой стороны, наиболее точное прогнозирование объёмов продаж нефтепродуктов позволит более эффективно использовать ресурсы страны, что не может не отразиться на экономике в целом. Поэтому, очень важно, чтобы компании, занимающиеся поставкой и продажей нефтепродуктов, имели представление о том, что может произойти в будущем, так как это позволит своевременно реагировать на возможные события. Кроме того, к процедуре прогнозирования объёма продаж следует подходить серьезно, поскольку из нее вытекает планирование всего бизнеса. Это может быть, как формирование бизнес-плана, так и формирование бюджета компании. Если прогноз является ошибочным и неточным, то и планирование окажется никому не нужным, так как не принесёт должных результатов.

Таким образом, планирование бизнеса, то есть формирование бизнес-планов, оценка и формирование бюджета, опирается на прогнозирование объема продаж, а целью планирования является распределение ресурсов компании таким образом, чтобы обеспечить эти ожидаемые продажи.

Компания может заниматься прогнозированием доли своих продаж на рынке, либо оценивать свой личный объем продаж. В данной работе будет рассматриваться прогнозирование личных объемов продаж нефтепродуктов компании ОАО «Газпром нефть». ОАО «Газпром нефть» предоставила статистику, по которой на данный момент, для того, чтобы рассчитать прогноз объемов продаж нефтепродуктов, а так же сопутствующих товаров на АЗС на следующий день тратится 4.5 человеко/дня. Такой показатель для компании является неприемлемым в виду отсутствия необходимого времени, нужного количества человеческих ресурсов, а так же в виду того, что процесс прогнозирования ведётся с помощью инструментов Excel, которые опираются на данные и факты, которые на сегодняшний момент потеряли свою актуальность и эффективность, что не позволяет строить удовлетворительный прогноз.

Поэтому можно выделить ряд прикладных задач, позволяющих упростить и модифицировать этот процесс. Одна из таких задач – это задача разработки ансамбля из разных моделей алгоритмов прогнозирования, который позволит выбрать лучший из возможных методов для каждого прогнозируемого объекта, сможет накапливать статистику о параметрах выбранной модели для дальнейшего анализа, а так же позволит сделать процесс создания прогноза более автоматизированным.

2. Терминология

В данной работе будут использоваться следующие понятия:

- Временной ряд - некоторая реализация процесса на равномерной сетке времени
- Сезонность - периодически повторяющаяся компонента временного ряда
- Тренд - общая систематическая линейная или нелинейная компонента временного ряда, которая может изменяться во времени
- MAPE - средняя абсолютная ошибка в процентах
- MPE - средняя ошибка в процентах
- SP - доля ошибки спрогнозированного значения от фактического в процентах

$$MAPE = \frac{1}{N} * \sum_{i=1}^N \frac{|X_i - \bar{X}_i|}{X_i} * 100\%$$

$$MPE = \frac{1}{N} * \sum_{i=1}^N \frac{(X_i - \bar{X}_i)}{X_i} * 100\%$$

$$SP = \frac{\sum_{i=1}^N \bar{X}_i}{\sum_{i=1}^N X_i} * 100\%$$

\bar{X}_i - спрогнозированное значение, X_i - фактическое значение, N – длина прогноза

3. Постановка задачи

Целью данной работы является расширение существующего в компании ОАО «Газпром нефть» кейса по прогнозированию объёма продаж сопутствующих товаров и нефтепродуктов на АЗС.

Самыми популярными марками бензинов в России являются АИ-92 и АИ-95. По данным на конец 2015 года 64% автовладельцев заправляются бензином 92-ой марки. Бензин АИ-95 предпочитают всего 19 %, однако этот показатель существенно выше в Москве и Санкт-Петербурге – 51%. Остальные виды топлива (АИ-80, ДТ, ДТ Евро и др.) пользуются меньшим спросом. Поэтому, в качестве исследуемых объектов были выделены следующие нефтепродукты:

1. Бензин Аи-92
2. Бензин Аи-95

В рамках данной работы были выделены следующие задачи:

- Разработать прогнозную модель алгоритмов прогнозирования, из которых будет выбираться лучший для произведения прогноза
Должен осуществляться ежедневный прогноз на следующие промежутки времени:
 1. на 1 день вперёд
 2. на 7 дней вперёд
 3. на 14 дней вперёд
 4. на 30 дней вперёд
- Протестировать получившийся ансамбль на интересующих тестовых данных
- Предложить возможные варианты по улучшению прогноза и точности

Отдел розницы ОАО «Газпром нефть» поставил так же дополнительную задачу, которая заключается в получении не только абсолютной процентной погрешности прогноза (МАРЕ), но и оценке суммарной доли ошибки прогнозируемого объёма

(SP) на 1 день вперёд, где “недобор” и “перебор” будут компенсировать друг друга внутри отделения. Получение данного показателя можно объяснить необходимостью знания того, в каких пределах будет находиться объём продаж завтрашнего дня и, соответственно, в каких пределах будет находиться выручка или прибыль внутри каждого отделения.

Автомобильно-заправочные станции компании ОАО “Газпром нефть” представлены на территории всей России, а так же на территории стран СНГ. В качестве единицы прогноза выбрана отдельная АЗС, но суммарная ошибка прогноза должна вычисляться внутри отделения.

АЗС компании “Газпром нефть” представлены в следующих отделениях:

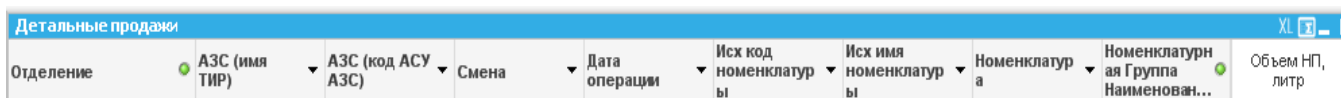
- Отделение Москва
- Газпромнефть – Азия
- Газпромнефть – Казахстан
- Газпромнефть – Таджикистан
- Газпромнефть – Юг
- Отделение Барнаул
- Отделение Екатеринбург
- Отделение Кемерово
- Отделение Красноярск
- Отделение Московская область
- Отделение Нижний Новгород
- Отделение Новосибирск
- Отделение Омск
- Отделение Санкт-Петербург
- Отделение Свердловская область
- Отделение Сургут
- Отделение Тюмень
- Отделение Челябинск
- Отделение Ярославль

В качестве прогнозируемых отделений были выбраны 2 наиболее значимых и больших отделения – Москва и Санкт-Петербург.

4. Описание и подготовка входных данных

- В качестве исследуемого промежутка времени были выбраны детальные продажи с 01.01.2015 по 31.03.2016
- В зависимости от выбранного алгоритма и его модели, обучение проводилось на промежутке с 01.01.2015 по 31.12.2015, тестирование проводилось на промежутке с 01.01.2016 по 31.03.2016
- В качестве интересующих отделений были выбраны отделение Москва (количество АЗС: 91) и отделение Санкт-Петербург (количество АЗС: 75)
- В качестве исследуемых нефтепродуктов были выбраны бензин АИ-92 и бензин АИ-95

Предоставленные данные по каждой АЗС хранятся в BI-системе QlikView [1] в следующем виде:



Отделение	АЗС (имя ТИР)	АЗС (код АСУ АЗС)	Смена	Дата операции	Исх код номенклатуры	Исх имя номенклатуры	Номенклатура	Номенклатурная Группа Наименован...	Объем НП, литр
-----------	---------------	-------------------	-------	---------------	----------------------	----------------------	--------------	-------------------------------------	----------------

Однако, для составления прогноза достаточно объёма, даты операции, уникального идентификатора АЗС(АЗС имя) и названия продукта (номенклатура).

При зрительном обзоре предоставленных данных было замечено, что данные не агрегированы по объёму нефтепродуктов, то есть могли присутствовать несколько записей о конкретной АЗС, конкретной дате и конкретному продукту, поэтому было проведено предварительное агрегирование данных.

Так же при составлении прогноза не учитывались АЗС у которых на интересующем промежутке времени было пропущенное значение хотя бы по одному дню, так как при ежедневном прогнозе, пропущенные данные могут сместить сезонность, тренд и другие важные показатели.

Для использования нейронных сетей была произведена нормализация данных.

Пример входных данных:

- АЗС (имя): 10_АЗС №1 Дунайский пр., д.29
- Дата: 01.01.2015
- Номенклатура: Бензин АИ-92
- Объём: 2127

5. Обзор

5.1. Подходы управления данными

В ноябре 2015 года компания M-Brain [2] провела исследование Market Intelligence Trends 2020 survey, результатом которого послужило появление 3-х основных трендов для систем управления данными до 2020 года.

1. Автоматизация сбора информации

Компании внедряют системы для сбора информации из множества источников практически в режиме реального времени, чтобы проактивно и быстрее конкурентов идентифицировать и находить рыночные возможности

Вызов: найти и внедрить лучшие инструменты для гибкой обработки разнообразных данных с использованием минимального числа интерфейсов.

Ключевые факторы успеха:

- Определить оптимальный набор инструментов данных и не допустить чрезмерного распыления активностей
- Оптимизировать расходы в целом на комплекс инструментов обработки и управления данными
- Гибкость решения для адаптации меняющимся бизнес-требованиям

2. Автоматизация анализа данных

С использованием инструментов Big Data и статической подготовки и проверки многие функции анализа данных будут полностью автоматизированы.

Вызовы: быть в курсе последних доступных возможностей автоматизации анализа, рассмотреть инструменты статистики и моделирования, найти партнёра - эксперта в этой теме

3. Кроссфункциональная интеграция

Автоматический сбор и анализ данных будут интегрированы с другими организационными функциями. Многие компании осознают ценность интеграции для обеспечения доступа ко всем данным через единую точку доступа.

Данная работа является исследованием, которое в дальнейшем может послужить продолжением в освоении подходов управления данными в ОАО “Газпром нефть”, такими как: определение оптимального набора инструментов данных, недопущение чрезмерного распыления активностей, оптимизация расходов и другими.

5.2. Методы прогнозирования

Методы прогнозирования объема продаж можно разделить на три основные группы:

1. методы экспертных оценок
2. методы анализа и прогнозирования временных рядов
3. казуальные (причинно-следственные) методы

В основе метода экспертных оценок лежит возможная перспектива развития и субъективная оценка текущего момента. Такие методы лучше всего показывают себя, когда невозможно точно получить непосредственную и достоверную информацию об интересующем явлении или процессе.

Методы анализа и прогнозирования временных рядов, а так же казуальные методы основываются на анализе количественных показателей. Метод анализа и прогнозирования временных рядов заключается в создании прогноза с некоторой вероятностью для случайной компоненты, которой, в свою очередь, соответствует некая детерминированная компонента, то есть значение, которой узнать не представляет особых трудностей, если известна её основная тенденция развития. Например, в качестве детерминированной компоненты может выступать время, а некоторый количественный показатель в этот момент времени нужно спрогнозировать.

Казуальные или причинно-следственные методы опираются на нахождение факторов, которые влияют на возможные значения прогнозируемого показателя. Для того чтобы найти эти факторы, необходимо построить модель экономического поведения объекта, которая будет учитывать и показывать взаимосвязи явлений и

процессов, связанные с рассматриваемым объектом.

Достоинства и недостатки присутствуют в каждой группе методов прогнозирования объёмов продаж. В данной работе будут рассматриваться методы анализа и прогнозирования временных рядов.

Прогнозирование можно разделить на краткосрочное или среднесрочное среднесрочные (день, неделя, месяц) и на долгосрочное (квартал, 1, 2 или более лет). Необходимо учитывать, что при прогнозировании на долгосрочный период, необходимо иметь большее количество данных, чем при краткосрочном периоде, так как при более длительном периоде больше факторов могут влиять на ожидаемый результат, соответственно необходимо выявить эти факторы при анализе имеющегося промежутка, которого может просто не хватить, из-за его краткосрочности. Как было сказано выше, в данной работе будет производиться прогнозирование на краткосрочный период – 1 день, 7 дней, 14 дней, 30 дней.

Как построить точный прогноз? Существует множество пунктов, на которые можно обратить внимание при построении прогноза объёма продаж, здесь представлены некоторые из них:

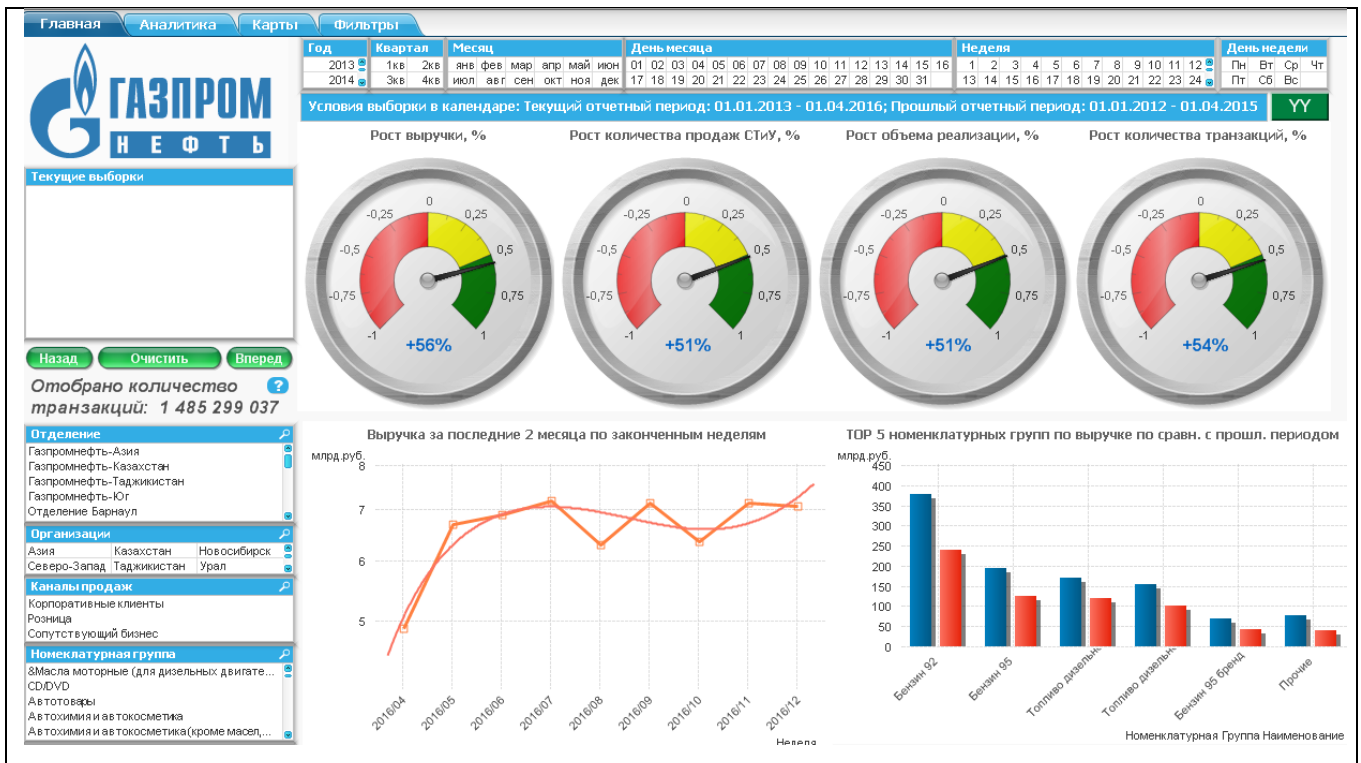
1. Достоверные данные о продажах за некоторый промежуток времени. В зависимости от того, на какой промежуток будет производиться прогноз, необходимо варьировать длину анализируемого промежутка. Если этот промежуток будет слишком мал, то будет не на что опираться при построении прогноза.
2. Коэффициент сезонности. Спрос на некоторые продукты может увеличиваться или уменьшаться в зависимости от времени года и т.д.
3. Эластичность спроса по цене на продукт. Этот показатель характеризует изменение спроса в зависимости от увеличения или падения цен.
4. Рост производственных мощностей. При появлении новых точек производства, таких как количество магазинов и т.п. необходимо отразить этот факт на прогнозе.
5. Коэффициент влияния внешних факторов (общеекономической ситуации в государстве и конкуренты).

5.3. Существующий кейс

На данном этапе в компании ОАО “Газпром нефть” реализован кейс по прогнозированию объёмов продаж сопутствующих товаров и нефтепродуктов на АЗС на один месяц вперёд.

Таким образом, создание ежедневного прогноза на более мелкие промежутки времени позволит получить более детальную картину для составления бизнес-стратегии и планирования.

Скриншоты существующего кейса, представление которых реализовано в Business intelligence - системе QlikView, позволяющей наглядно оценить прогноз объёма продаж сопутствующих товаров и нефтепродуктов на АЗС на месяц вперёд, представлены ниже:



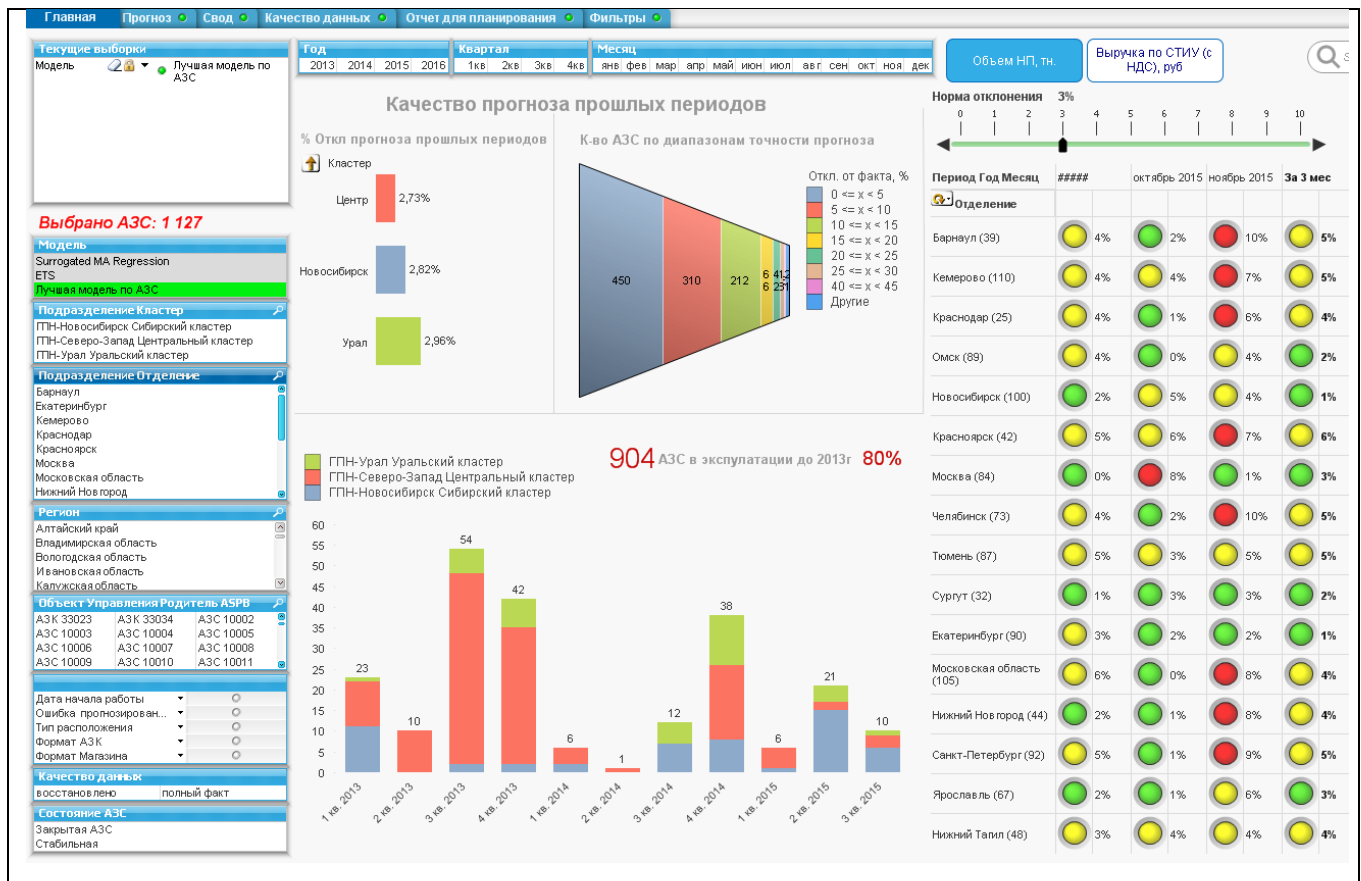


Рис. 1. Кейс ОАО “Газпром нефть” для прогнозирования на месяц вперед.

5.4. Связанные работы

Алгоритмы прогнозирования применялись к области составления прогноза объёма продаж и ранее. В этих исследовательских работах рассматривались различные методы, такие как: ARIMA – модели, методы экспоненциального сглаживания, метод Хольта-Винтерса [3], [4], [5], [6], [7], нейронные сети [8], [9], [10]. Однако многие из существующих исследовательских работ посвящены одному из методов прогнозирования и не рассматривают возможности создания ансамбля нескольких методов, из которых будет выбираться наилучшая модель. Ансамблирование методов позволит делать более точный прогноз, так как для различных рассматриваемых объектов, лучший результат могут показывать различные модели. Также стоит отметить, что данные работы не подразумевают возможность накопления статистики параметров лучшей модели для рассматриваемых объек-

тов, что существенно может облегчить задачу создания прогноза в случае потери или отсутствия входных данных. В качестве вспомогательных пакетов для исследования использовались следующие пакеты языка R: `neuralnet` [11] и `forecast` [12].

6. Используемые алгоритмы

В качестве используемых методов прогнозирования были рассмотрены следующие алгоритмы:

6.1. Holt-Winters

В методе Хольта используются три расчетных формулы:

1. Экспоненциально сглаженный ряд (оценка текущего уровня)

$$L_t = ay_t + (1 - a)(L_{t-1} - T_{t-1})$$

2. Оценка тренда

$$T_t = b(L_t - L_{t-1}) + (1 + b)T_{t-1}$$

3. Прогноз на p периодов вперед

$$\bar{Y}_t = L_t + pT_t,$$

где a, b — постоянные сглаживания из интервала $[0, 1]$.

Метод Винтерса совершенствует метод Хольта [15], вводя в рассмотрение уравнение, которое учитывает сезонность:

1. Экспоненциально сглаженный ряд

$$a \frac{Y_t}{S_{t-s}} + (1 - a)(L_{t-1} + T_{t-1})$$

2. Оценка тренда

$$T_t = b(L_t - L_{t-1}) + (1-b)T_{t-1}$$

3. Оценка сезонности

$$S_t = y \frac{Y_t}{L_t} + (1-y)S_{t-s}$$

4. Прогноз на p периодов вперед

$$\bar{Y}_{t+p} = (L_t + pT_t)S_{t-s+p},$$

где a, b, y — постоянные сглаживания для уровня, тренда и сезонности, соответственно; S - длительность периода сезонного колебания. Оптимальные параметры подбираются путем минимизации среднеквадратичной ошибки прогноза.

6.2. Arima

Модель Arima(p,d,q) для нестационарного временного ряда X_t имеет вид:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j e_{t-j} + e_t$$

где e_t - стационарный временной ряд; c, a_i, b_j - параметры модели.

Δ^d - оператор разности временного ряда порядка d (последовательное взятие d раз разностей первого порядка - сначала от временного ряда, затем от полученных разностей первого порядка, затем от второго порядка и т.д.)

В рассматриваемой задаче был выбран метод autoArima, где параметры модели подбираются автоматически.

6.3. ETS

Метод экспоненциального сглаживания, позволяющий учитывать характер ошибки, тренда и сезонности. Формулы для расчёта каждого из вариантов представлены ниже:

Трендовая компонента	Сезонная компонента		
	N	A	M
N	$\hat{y}_{t+h} = l_t$ $l_t = \alpha y_t + (1-\alpha)l_{t-1}$	$\hat{y}_{t+h} = l_t - c_{t-s+h}$ $l_t = \alpha(y_t - c_{t-s}) + (1-\alpha)l_{t-1}$ $c_t = \gamma(y_t - l_{t-1}) + (1-\gamma)c_{t-s}$	$\hat{y}_{t+h} = l_t c_{t-s+h}$ $l_t = \alpha(y_t/c_{t-s}) + (1-\alpha)l_{t-1}$ $c_t = \gamma(y_t/l_{t-1}) + (1-\gamma)c_{t-s}$
A	$\hat{y}_{t+h} = l_t + h \cdot b_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h} = l_t - h b_t - c_{t-s+h}$ $l_t = \alpha(y_t - c_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$ $c_t = \gamma(y_t - (l_{t-1} + b_{t-1})) + (1-\gamma)c_{t-s}$	$\hat{y}_{t+h} = (l_t - h b_t) c_{t-s+h}$ $l_t = \alpha(y_t/c_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)b_{t-1}$ $c_t = \gamma(y_t/(l_{t-1} + b_{t-1})) + (1-\gamma)c_{t-s}$
Ad	$\hat{y}_{t+h} = l_t + \sum_{\tau=1}^h \varphi^\tau b_t$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} + \varphi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\varphi b_{t-1}$	$\hat{y}_{t+h} = l_t + \sum_{\tau=1}^h \varphi^\tau b_t - c_{t-s+h}$ $l_t = \alpha(y_t - c_{t-s}) + (1-\alpha)(l_{t-1} + \varphi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\varphi b_{t-1}$ $c_t = \gamma(y_t - (l_{t-1} + \varphi b_{t-1})) + (1-\gamma)c_{t-s}$	$\hat{y}_{t+h} = \left(l_t + \sum_{\tau=1}^h \varphi^\tau b_t \right) c_{t-s+h}$ $l_t = \alpha(y_t/c_{t-s}) + (1-\alpha)(l_{t-1} + \varphi b_{t-1})$ $b_t = \beta^*(l_t - l_{t-1}) + (1-\beta^*)\varphi b_{t-1}$ $c_t = \gamma(y_t/(l_{t-1} + \varphi b_{t-1})) + (1-\gamma)c_{t-s}$
M	$\hat{y}_{t+h} = l_t b_t^h$ $l_t = \alpha y_{t-1} + (1-\alpha)(l_{t-1} b_{t-1})$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}$	$\hat{y}_{t+h} = l_t b_t^h - c_{t-s+h}$ $l_t = \alpha(y_t - c_{t-s}) + (1-\alpha)(l_{t-1} b_{t-1})$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}$ $c_t = \gamma(y_t - (l_{t-1} b_{t-1})) + (1-\gamma)c_{t-s}$	$\hat{y}_{t+h} = l_t b_t^h c_{t-s+h}$ $l_t = \alpha(y_t/c_{t-s}) + (1-\alpha)(l_{t-1} b_{t-1})$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}$ $c_t = \gamma(y_t/(l_{t-1} b_{t-1})) + (1-\gamma)c_{t-s}$
Md	$\hat{y}_{t+h} = l_t b_t^{\sum_{\tau=1}^h \varphi^\tau}$ $l_t = \alpha y_t + (1-\alpha)(l_{t-1} b_{t-1}^\varphi)$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}^\varphi$	$\hat{y}_{t+h} = l_t b_t^{\sum_{\tau=1}^h \varphi^\tau} - c_{t-s+h}$ $l_t = \alpha(y_t - c_{t-s}) + (1-\alpha)(l_{t-1} b_{t-1}^\varphi)$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}^\varphi$ $c_t = \gamma(y_t - (l_{t-1} b_{t-1}^\varphi)) + (1-\gamma)c_{t-s}$	$\hat{y}_{t+h} = l_t b_t^{\sum_{\tau=1}^h \varphi^\tau} c_{t-s+h}$ $l_t = \alpha(y_t/c_{t-s}) + (1-\alpha)(l_{t-1} b_{t-1}^\varphi)$ $b_t = \beta^*(l_t/l_{t-1}) + (1-\beta^*)b_{t-1}^\varphi$ $c_t = \gamma(y_t/(l_{t-1} b_{t-1}^\varphi)) + (1-\gamma)c_{t-s}$

Рис. 2. Формулы для вычисления прогноза по методу ETS.

- N – компонента отсутствует;
- A – компонента носит аддитивный характер;
- M – компонента носит мультипликативный характер;
- Ad – компонента носит аддитивный характер демпфированного тренда;
- Md – компонента носит мультипликативный характер демпфированного тренда.

6.4. Нейронные сети

В качестве архитектуры нейронной сети выбран перцептрон [13] с одним скрытым слоём, количество нейронов в скрытом слое является изменяемым параметром. Для обучения используется алгоритм обратного распространения ошибки [14].

Обучение нейронной сети проводилось на 80% данных: промежуток с 01.01.2015 по 31.12.2015, тестирование проводилось на 20% данных: промежуток с 01.01.2016 по 31.03.2016.

Для обучения и тестирования нейронных сетей было выбрано несколько моделей, принимающих в качестве входного сигнала различные варианты данных:

Прогнозирование на 1 день вперёд

- по 3-м предыдущим дням
- по 7-ми предыдущим дням
- по 14-ти предыдущим дням
- по 7-ми предыдущим дням через 7 (т.е. если прогноз делается на понедельник, т.е. в качестве входных данных берутся последние 7 понедельников)

Прогнозирование на 7 дней вперёд

- по 14-ти предыдущим дням

7. Ансамблирование и накопление статистики

Каждый из рассмотренных алгоритмов содержит в себе ещё несколько вариантов, которые обусловлены следующими показателями:

- длина моделируемого промежутка
- длина сезонного цикла
 - 7 – недельный
 - 14 – двухнедельный
 - 30 – месячный

- 365 - годовой

Для накопления и анализа статистики в дальнейшем, для каждой АЗС и для каждого рассмотренного отделения сохраняются данные о следующих параметрах:

- лучший алгоритм
- лучшая длина моделируемая промежутка
- лучшее количество скрытых нейронов
- лучшая длина сезонного цикла

Понятие лучшего параметра основано на оценке абсолютной средней процентной ошибки - MAPE.

Таким образом, сохраняются все параметры модели, которая достигла наименьшей ошибки MAPE на тестируемых данных.

Все используемые модели объединены в ансамбль алгоритмов. Для каждого интересующего промежутка прогноза, для каждой АЗС в отдельности выбирается наилучшая модель, и впоследствии, эта модель будет использоваться для составления прогноза.

Накопление статистики по-большой части производится для экспертов в данной области, которые смогут использовать полученные данные для составления прогноза в том случае, когда автоматически это невозможно. То есть если прогноз не может быть построен в виду отсутствия нужного количества данных или пропуска значений. В таких случаях прогноз должен осуществляться человеком с использованием параметров, которые лучше всего себя показали на “похожих” АЗС. Понятие “похожести” АЗС может быть связано с такими параметрами как: формат АЗС, месторасположение, новизна и другие.

8. Скользящий контроль

В задаче прогнозирования, где главным рассматриваемым объектом является временной ряд, важное значение имеет то, что значения ряда упорядочены во

времени, поэтому вариантов скользящего контроля не так много. В данной работе, за счёт большого количества параметров в каждой модели алгоритма, учитывались следующие варианты скользящего контроля:

1. Контроль при нарастающей длине обучения

В алгоритмах Holt-Winters, ETS и ARIMA изменяемым параметром является длина моделируемого промежутка, поэтому из нескольких длин обучения выбирается та, которая показала наименьшую ошибку.

2. Контроль при фиксированной длине обучения

В этом случае фиксируется длина моделируемого промежутка, который “скользит” по тестовой выборке. За счёт этого итоговая погрешность считается как среднее арифметическое получившихся ошибок. Данный подход не позволяет модели адаптироваться под конкретное тестовое множество.

9. Эксперименты

В представленных таблицах показаны результаты проведённых экспериментов на основании построенного ансамбля моделей алгоритмов прогнозирования:

9.1. Holt-Winters

- Прогноз на 1 день вперёд

Отделение	Бензин АИ-92 MAPE / SP, %	Бензин АИ-95 MAPE / SP, %
Москва	1.95 / 2.68	1.86 / 2.50
Санкт-Петербург	1.53 / 2.45	1.57 / 2.28

- Прогноз на 7, 14, 30 дней:

Бензин АИ-92

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	7.12 / 3.06	12.12 / 8.19	8.15 / 3.18
Санкт-Петербург	6.58 / 3.02	9.27 / 5.75	7.77 / 3.62

Бензин АИ-95

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	7.59 / 4.78	13.84 / 9.34	8.48 / 3.64
Санкт-Петербург	6.98 / 3.15	11.61 / 8.38	8.41 / 4.17

9.2. Arima

- Прогноз на 1 день вперёд

Отделение	Бензин АИ-92 MAPE / SP, %	Бензин АИ-95 MAPE / SP, %
Москва	2.13 / 3.53	2.25 / 3.94
Санкт-Петербург	1.45 / 2.12	1.98 / 2.76

- Прогноз на 7, 14, 30 дней:

Бензин АИ-92

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	9.24 / 5.22	10.39 / 5.90	10.68 / 4.42
Санкт-Петербург	9.67 / 3.21	8.15 / 5.21	9.54 / 3.42

Бензин АИ-95

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	9.77 / 6.16	10.37 / 5.66	13.65 / 5.76
Санкт-Петербург	9.83 / 3.58	10.78 / 3.94	11.39 / 3.90

9.3. ETS

- Прогноз на 1 день вперёд

Отделение	Бензин АИ-92 MAPE / SP, %	Бензин АИ-95 MAPE / SP, %
Москва	2.67 / 2.61	2.62 / 3.72
Санкт-Петербург	1.95 / 2.37	2.16 / 2.83

- Прогноз на 7, 14, 30 дней:

Бензин АИ-92

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	9.06 / 5.31	10.66 / 6.53	11.81 / 5.78
Санкт-Петербург	8.00 / 4.70	9.58 / 4.61	11.32 / 4.11

Бензин АИ-95

Отделение	7 дней, MAPE / MPE, %	14 дней, MAPE / MPE, %	30 дней, MAPE / MPE, %
Москва	9.38 / 5.88	10.15 / 5.58	11.45 / 5.78
Санкт-Петербург	8.28 / 5.47	9.20 / 4.57	11.19 / 4.51

9.4. Нейронные сети

Прогнозирование на 1 день вперёд

- по 3-м предыдущим дням (1)
- по 7-ми предыдущим дням (2)
- по 14-ти предыдущим дням (3)
- по 7-ми предыдущим дням через 7 (4)

Прогнозирование на 7 дней вперёд

- по 14-ти предыдущим дням (5)

Отделение	(1) MAPE/MPE/SP, %	(2) MAPE/MPE/SP, %	(3) MAPE/MPE/SP, %	(4) MAPE/MPE/SP, %	(5) MAPE/MPE, %
Москва	11.63 / 2.03 / 3.17	9.93 / 1.71 / 3.03	11.80/2.02 / 4.37	9.34 / 3.14 / 2.91	10.38 / 5.32
Санкт-Петербург	11.24 / 1.53 / 3.21	10.05 / 2.23 / 3.47	11.11 /2.49 /3.11	8.88 / 2.87 / 3.02	9.12 / 4.67

- 95 бензин

Отделение	(1) MAPE/MPE/SP, %	(2) MAPE/MPE/SP, %	(3) MAPE/MPE/SP, %	(4) MAPE/MPE/SP, %	(5) MAPE/MPE , %
Москва	12.00 / 1.80 / 4.22	12.50 / 1.57 / 3.32	11.25 / 1.77 / 4.31	11.47 / 0.88 / 3.17	11.60 / 5.72
Санкт-Петербург	11.58 / 2.76 / 3.91	11.03 / 2.39 / 3.12	11.21 / 2.15 / 4.44	10.65 / 1.98 / 1.65	9.74 / 4.21

Графики некоторых лучших моделей представлены в приложении к данной работе.

10. Результаты

Для того чтобы более наглядно можно было оценить результаты работы моделей на тестовых данных, полученные результаты были усреднены по очереди по отделению, длине прогнозируемого промежутка и прогнозируемому объекту. В качестве основного параметра оценивания работы модели выбрана средняя абсолютная процентная ошибка (MAPE):

1. Усреднен бензин и длина прогноза, сравнивается отделение.

	Москва, %	Санкт-Петербург, %
Holt-Winters	7.64	6.72
ETS	8.48	7.71
ARIMA	8.56	7.85
Нейронные сети	11.19	10.46

2. Усреднено отделение и длина прогноза, сравнивается бензин.

	Бензин АИ-92, %	Бензин АИ-95, %
Holt-Winters	6.84	7.51
ETS	8.12	8.09
ARIMA	7.73	8.75
Нейронные сети	10.35	11.33

3. Усреднено отделение и бензин, сравнивается длина прогноза.

	1 день, %	7 дней, %	14 дней, %	30 дней, %
Holt-Winters	1.73	7.07	11.71	8.20
ETS	2.25	8.68	9.89	11.44
ARIMA	1.95	9.63	9.92	11.32
Нейронные сети	10.98	10.21	–	–

На основании полученных результатов сделаны следующие выводы:

- Бензин АИ-92 прогнозируется лучше Бензина АИ-95
- Отделение Санкт-Петербург прогнозируется лучше отделения Москва
- Holt-Winters показал наилучшие результаты, нейронные сети показали себя значительно хуже других алгоритмов
- Лучше всего строится прогноз на 1 день вперёд

11. Заключение

В ходе данной работы была рассмотрена задача прогнозирования объёмов продаж основных нефтепродуктов на АЗС компании ОАО "Газпром нефть". Были применены некоторые способы преобразования данных, была рассмотрена возможность применения различных моделей алгоритмов прогнозирования на реальных тестовых данных. Были проведены сравнение и анализ результатов.

Рассмотренные модели алгоритмов были объединены в единый ансамбль, реализованный на языке R, который в дальнейшем будет расширяться и использоваться сотрудниками компании ОАО "Газпром нефть".

Разработанный прогнозный модуль позволяет накапливать статистику о параметрах моделей алгоритмов, показавших наилучшие результаты. Существование таких параметров позволит экспертам по анализу данных в данной области, производить прогноз самостоятельно, когда это становится невозможным в виду отсутствия качества или нужного количества данных.

По результатам проведённых экспериментов лучшие результаты показал метод Хольта-Винтерса, усреднённая погрешность в этом случае составила 7.2%, хуже всего отработала нейронная сеть, здесь усреднённая погрешность оказалась равной 10.8%.

В ходе дальнейшего исследования можно рассмотреть другие алгоритмы прогнозирования, что может положительно сказаться на качестве прогноза и повысить точность. Кроме того, в данных не учитываются возможные простои АЗС,

учёт которых позволил бы не исключать из рассмотрения временные ряды с пропущенными значениями.

Так же, для улучшения качества прогноза в будущем можно учитывать факт изменения формата АЗС, который зависит от размера автозаправочной станции; возможно введение меток для входных данных, таких как: “праздничный день”, “пятница” и др., что позволит сделать дополнительный акцент на характере временного ряда.

12. Список литературы

1. Qlik: Business Intelligence | Data Visualization Tools URL: <http://www.qlik.com> (дата обращения: 01.03.2016)
2. Market Intelligence - Dobney.com URL: http://www.dobney.com/market_intelligence.htm (дата обращения: 12.03.2016)
3. Chatfield, C, (2004). The analysis of time series: An introduction, 6th edition, Boca Raton: Chapman & Hall/CRC Press.
4. Holt, C.C. (2004a). Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting, 20, 5-10.
5. Nau Robert, (2015), Statistical forecasting: notes on regression and time series analysis
6. Prajakta S. Kalekar Time series Forecasting using Holt-Winters Exponential Smoothing // Kanwal Rekhi School of Information Technology, 2006.
7. Rob J. Hyndman, Anne B. Koehler, J. Keith Ord, Ralph Snyder, (2008). Forecasting with Exponential Smoothing, The State Space Approach, Springer
8. Создаем нейронную сеть для прогнозирования временного ряда // Математическое бюро, Прогнозирование на ОРЭМ URL: <http://www.mbureau.ru/blog/sozdaem-neyronnuyu-set-dlya-prognozirovaniya-vremennogo-ryada> (дата обращения: 12.03.2016).
9. R.J.Frank, N.Davey, S.P.Hunt Time Series Prediction and Neural Networks. Department of Computer Science, University of Hertfordshire, Hatfield, UK.
10. Dorffner, G. 1996, Neural Networks for Time Series Processing. Neural Network World 4/96, 447-468.
11. Package „neuralnet“ // cran.r-project.org URL: <https://cran.rproject.org/web/packages/neuralnet/neuralnet.pdf> (дата обращения: 10.02.2016).
12. Package „forecast“ // cran.r-project.org URL: <https://cran.rproject.org/web/packages/forecast/forecast.pdf> (дата обращения: 02.02.2016)

13. Персептрон URL:

<http://www.machinelearning.ru/wiki/index.php?title=Персептрон> (дата обращения: 13.04.2016)

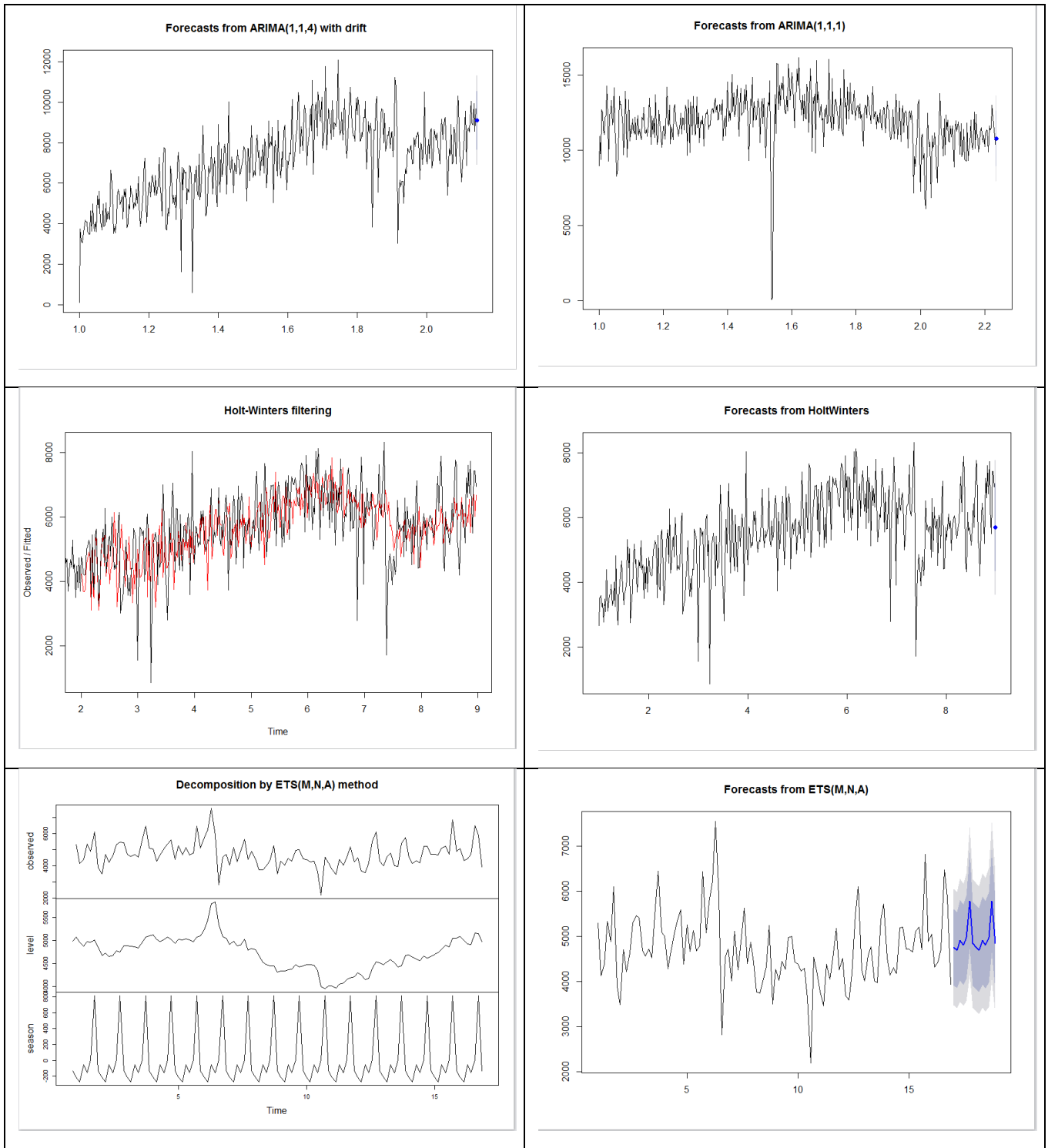
14. Алгоритм обратного распространения ошибки URL:

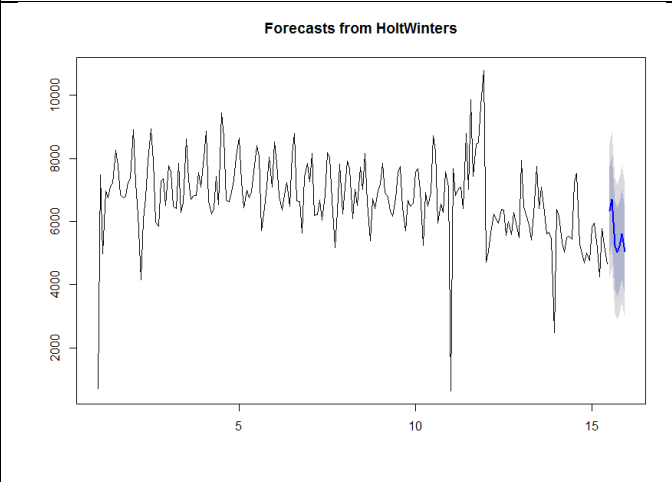
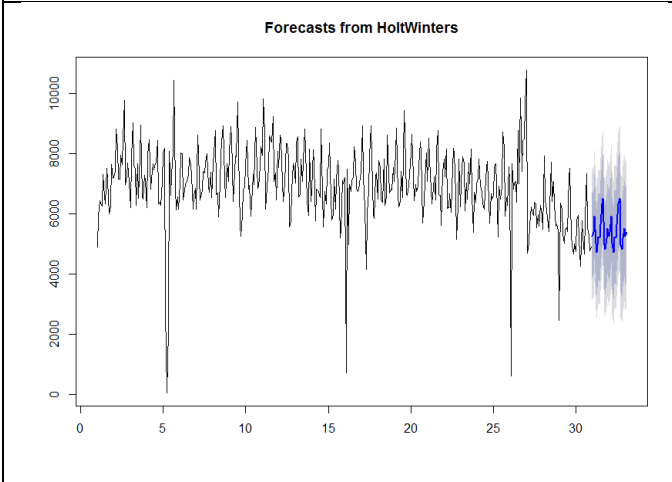
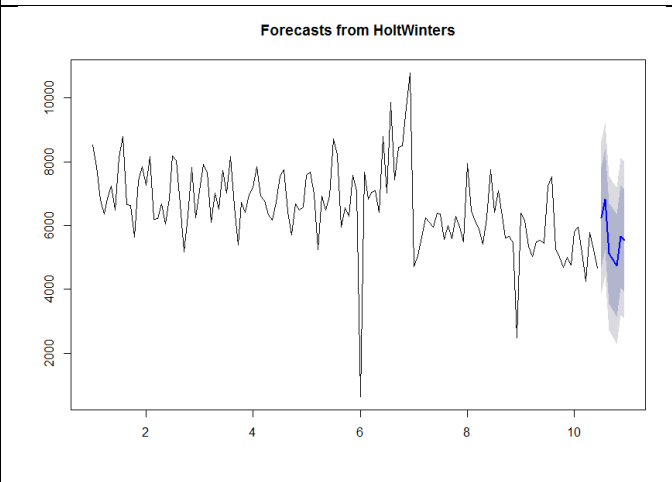
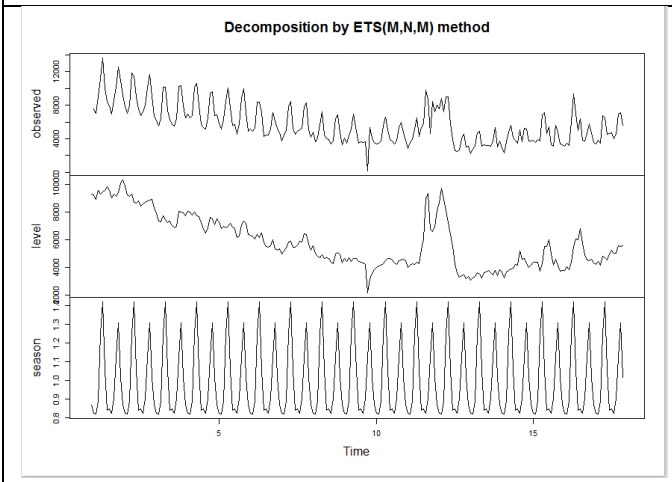
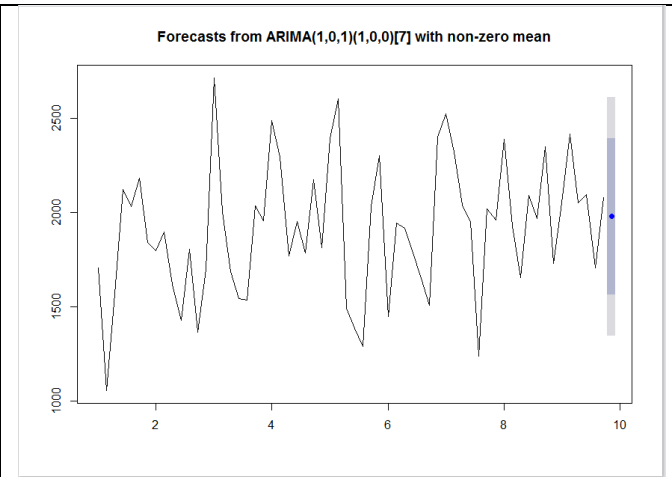
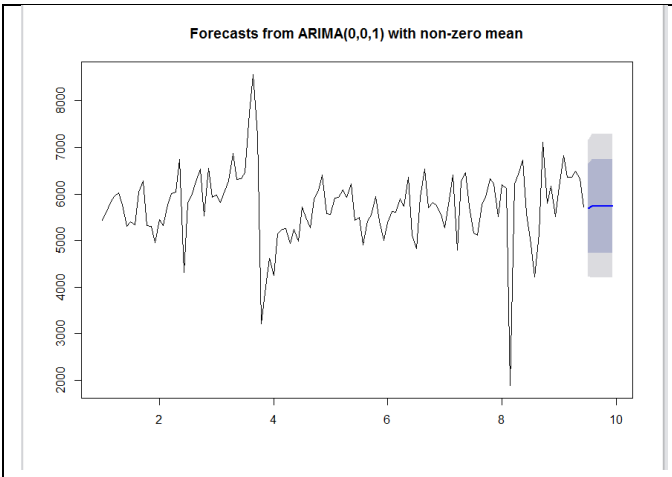
<https://habrahabr.ru/post/154369/> (дата обращения: 13.04.2016)

15. Воронцов К.В. Лекции по прогнозированию временных рядов URL:

<http://www.machinelearning.ru/wiki/images/c/cb/Voron-ML-forecasting-slides.pdf>
(дата обращения: 03.02.2016).

13. Приложение





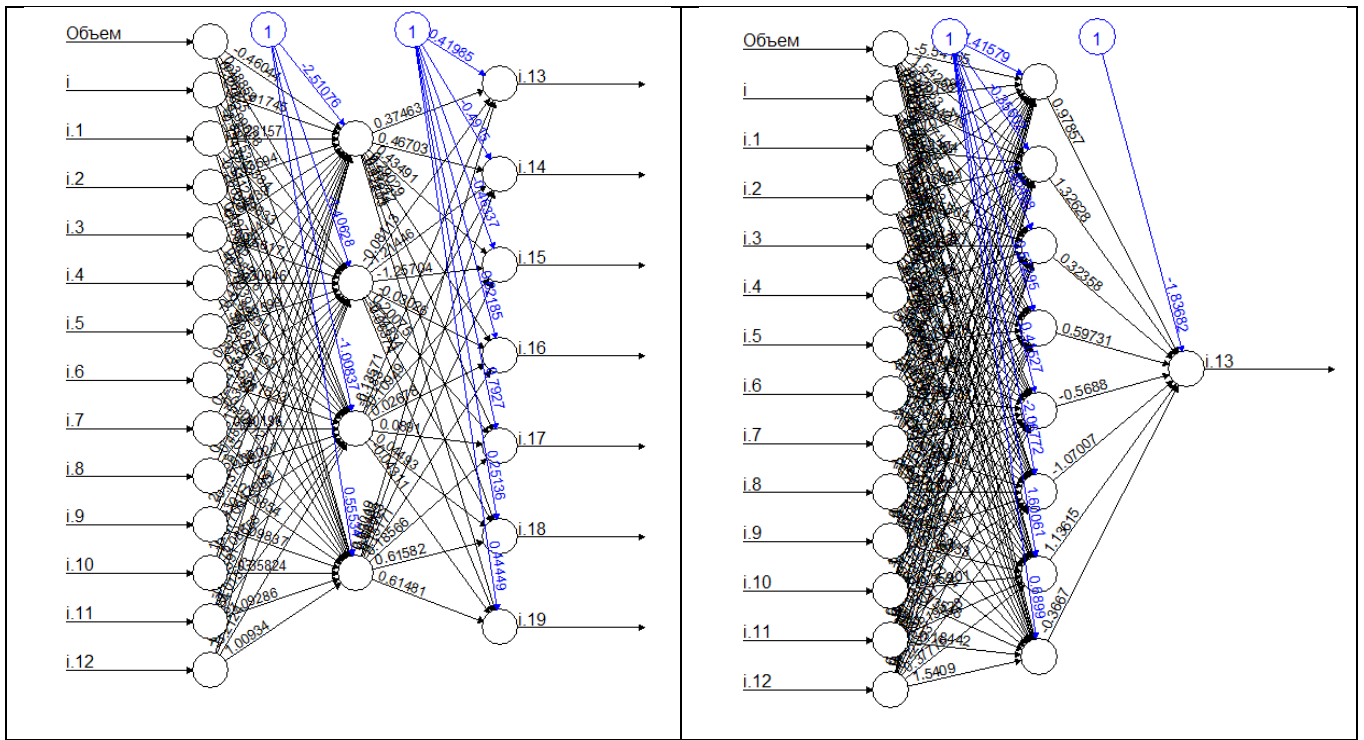


Рис. 2. Реальные тестовые данные. Обучение прогнозных моделей.