

Оглавление

Проблема исследования	1
Актуальность её решения	2
Основная идея реализации	3
Конкретизация реализации	5
Заключение	17
Список используемых источников	17
Примечания	18

Проблема исследования

В данном проекте рассматривается тема использования методов статистики и машинного обучения с целью получения точных прогнозных планов в разных временных разрезах на основе процедуры прогнозирования временных рядов. Отличительными чертами проблемы, поднятой в рамках данной работы, являются наличие нескольких категорий прогнозируемых явлений с индивидуальными поведением рядов, что побуждает подбирать отдельные решения для каждого отдельного случая, а также отсутствие каких-либо вторичных признаков и характеристик, которые могли бы быть использованы для выполнения процедуры прогнозирования, кроме численных величин наблюдений. В связи с вышеупомянутыми спецификами следует опираться на множество методов и даже, возможно, проводить их комбинацию при условии оптимальности этой тактики с целью диверсификации инструментов для решения задач разного рода.

Необходимо отметить, что набор данных, над которым будет проводиться работа исходит из реальной коммерческой деятельности. Этот факт будет доказательством эффективности разрабатываемого решения при условии успешности выполнения им заявленных задач.

Помимо практической части, в рамках выполнения текущего проекта будут разобраны механики работы некоторых методов прогнозирования, а также лично разработанных инструментов, которые подтвердили свою продуктивность в

рамках профессиональной деятельности. Более того, будет составлен обзор результатов с формированием выводов, касаемых успешности разработанного решения и набора конкурентных преимуществ и прочих выгод, которые он способен предоставить.

Актуальность её решения

Современный мир бизнеса представляет из себя комплексную и динамичную среду, которой присущи существование огромных массивов неструктурированных данных, использование и последующая обработка которых позволяют получить доступ к возможностям в разработке планов и стратегий деятельности предприятия для разных временных отрезков, великое множество взаимозависимых и параллельно протекающих бизнес-процессов, происходящих как в пределах фокусной компании, так и вне её, особо завышенные требования со стороны потребителей и высокий уровень конкуренции на рынке, заставляющий его игроков постоянно совершенствоваться с целью выживания и удержания лидерских позиций по отношению к своим соперникам. Таким образом, чтобы заранее адаптироваться к быстроизменяющейся внешней среде предпринимательской деятельности, тем самым сделав первые шаги к получению конкурентного преимущества на рынке, многие организации прибегают к планированию. Именно оно позволяет добиваться гибкости функционирования предприятия, интегрировать протекающие в нём взаимосвязанные бизнес-процессы, более грамотно распределять ресурсы, предоставлять качественный сервис потребителям с оптимальным уровнем издержек, а также устранять риск возникновения межфункциональных и межорганизационных конфликтов. Планирование играет ключевую роль в современном бизнесе, так как оно предоставляет возможность в адаптации к текущим экономическим трендам, требующих высокую эффективность функционирования компании в условиях неопределённого будущего.

В связи с вышеизложенным можно сделать вывод о том, что эффективная система прогнозирования будет служить существенным конкурентным преимуществом для современной коммерческой компании. Именно разрабатываемое ИТ-решение может быть предназначено для решения любых задач, обладающих схожей спецификой, независимо от сектора, в котором оперирует целевое предприятие, предоставляющее свои исторические данные для последующей их последующей обработки.

Основная идея реализации

Разрабатываемое ИТ-решение строится на чёткой последовательности действий и представляет из себя циклический алгоритм действий, который функционирует, пока существуют какие-либо временные ряды датасета, которые необходимо спрогнозировать. Таким образом, последовательность команд алгоритма можно представить следующей таблицей (*Таблица 1*):

Таблица 1. Алгоритм действий

№ Шага	Описание действия
1	Установка временного окна прогнозирования пользователем;
2	Считывание датасета и его базовая предобработка, заключающаяся в удалении строк с пустыми значениями, преобразования столбца с датами и последующей сортировкой значений в нём;
3	Проведение базового разведочного анализа временных рядов, содержащихся в датасете, с помощью визуальных инструментов аналитики с целью ознакомления со спецификой материала, над которым будет вестись работа;
3.1	Визуализация факта, месячной скользящей с трендом для каждого временного ряда датасета;
3.2	Визуализация декомпозиции компонент для каждого временного ряда датасета;
3.3	Визуализация боксплотов (ящиков с усами) для анализа выбросов для каждого временного ряда датасета;

3.4	Визуализация распределений значений фактов для каждого временного ряда датасета;
4	Формирование функции вызова и настройка параметров модели класса «deep learning», выражающейся в нейронной сети, для последующего её применения при прогнозировании;
5	Проведение процедуры прогнозирования временных рядов в датасете;
5.1	Преобразование временных разрезов временных рядов датасета из дневных в месячные в формате «ГГГГ-ММ»;
5.2	Применение личной технологии для получения дневных коэффициентов для дневной декомпозиции первого прогнозируемого месяца заданного временного окна для каждого временного ряда датасета;
5.3	Разбитие каждого временного ряда датасета на обучающую и тестовую выборки (20% конечных наблюдений – это тестовый период) с целью последующей проверки эффективности каждого инструмента из набора выбранных моделей;
5.4	Проведение процедуры ex-post прогнозирования с помощью каждого инструмента из набора выбранных моделей для каждого временного ряда датасета;
5.5	Расчёт набора выбранных ошибок прогнозирования на тестовых периодах каждого временного ряда датасета каждого использованного инструмента прогнозирования из набора выбранных моделей;
5.6	Отбор 2-ух лучших инструментов прогнозирования на тестовом периоде каждого временного ряда датасета по значениям средних абсолютных ошибок в процентах (MAPE) и их последующая комбинация посредством перемножения присвоенных им коэффициентов, нормированных по величинам их квадратичных ошибок (RMSE) соответственно, на величины их прогнозных планов и сложением полученных значений в прогнозные планы комбинационных моделей для каждого временного ряда датасета;
5.7	Расчёт ошибок прогнозирования полученных комбинационных моделей каждого временного ряда датасета на их тестовых периодах;
5.8	Отбор оптимального отдельного инструмента прогнозирования из установленного набора или их комбинационной модели для каждого временного ряда датасета согласно показанным им величиной средней абсолютной ошибки в процентах (MAPE) на его тестовом периоде;
5.9	Обучение присвоенных каждому временному ряду датасета его оптимальной модели прогнозирования на всём его периоде;

5.10	Осуществление процедуры прогнозирования динамик временных рядов датасета с помощью присвоенной каждому из них оптимальной модели на временной период, указанный пользователем ранее;
6	Вывод результатов работы построенных моделей прогнозирования для каждого временного ряда датасета;
6.1	Вывод таблицы ошибок прогнозирования каждого инструмента прогнозирования из выбранного набора для тестовых периодов каждого временного ряда датасета: Колонки: «Вид ошибки» - «Название временного ряда (категория)»; Строки: «Инструмент прогнозирования»
6.2	Вывод таблицы распределения оптимальных моделей прогнозирования для каждого временного ряда датасета по принципу «Категория» - «Оптимальная модель»;
6.3	Вывод визуализаций графиков факта и прогноза, построенного оптимальной моделью прогнозирования для каждого временного ряда датасета;
6.4	Вывод таблицы с месячным прогнозным планом для каждого временного ряда датасета со столбцом прогнозируемых дат, которые были обусловлены временным окном, выбранным пользователем ранее;
6.5	Вывод таблицы с дневной декомпозицией первого месяца прогнозного плана каждого временного ряда датасета согласно ранее рассчитанным дневным коэффициентам;
7	Вывод полного времени работы алгоритма.

Конкретизация реализации

1) Входные данные:

Реализация решения начинается с вводом пользователем временного окна, в течение которого временные ряды будут прогнозироваться. Фрагмент области для ввода данных выглядит следующим образом (*Рисунок 1*):

```
#Установка периода прогнозирования:
from datetime import datetime
start_time = datetime.now()
print('\033[1m' + 'УСТАНОВКА ПЕРИОДА ПРОГНОЗИРОВАНИЯ:' + '\033[0m')
cd1 = 'Начальный месяц для месячного прогнозирования и его дневной детализации (в формате "ГГГГ-ММ)": '
cd2 = 'Конечный месяц для месячного прогнозирования (в формате "ГГГГ-ММ)": '
daily = input(cd1)
monthly = input(cd2)
inp = 'salesdaily.csv' # Путь к файлу с историческим фактом
```

УСТАНОВКА ПЕРИОДА ПРОГНОЗИРОВАНИЯ:
Начальный месяц для месячного прогнозирования и его дневной детализации (в формате "ГГГГ-ММ"): 2019-08
Конечный месяц для месячного прогнозирования (в формате "ГГГГ-ММ"):

Рисунок 1. Фрагмент окна для установки временного окна прогнозирования

В качестве примера датасета, содержащего временные ряды для выполнения процедуры прогнозирования, была взята таблица ежедневных продаж разных категорий лекарственных препаратов в массовом выражении, зафиксированных в 600000 транзакциях за 6-летний период (с 2014 по 2019) в реальной торговой фармацевтической точке [3]. Эта таблица представлена в формате «.csv», обладает 2107 строками и 13 столбцами. Фрагмент считанной таблицы выглядит следующим образом (*Рисунок 2*):

```
dynam = pd.read_csv(inp).dropna(how='all')
dynam
```

	datum	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06	Year	Month	Hour	Weekday Name
0	1/2/2014	0.00	3.670	3.40	32.40	7.0	0.0	0.0	2.00	2014	1	248	Thursday
1	1/3/2014	8.00	4.000	4.40	50.60	16.0	0.0	20.0	4.00	2014	1	276	Friday
2	1/4/2014	2.00	1.000	6.50	61.85	10.0	0.0	9.0	1.00	2014	1	276	Saturday
3	1/5/2014	4.00	3.000	7.00	41.10	8.0	0.0	3.0	0.00	2014	1	276	Sunday
4	1/6/2014	5.00	1.000	4.50	21.70	16.0	2.0	6.0	2.00	2014	1	276	Monday
...
2101	10/4/2019	7.34	5.683	2.25	22.45	13.0	0.0	1.0	1.00	2019	10	276	Friday
2102	10/5/2019	3.84	5.010	6.00	25.40	7.0	0.0	0.0	0.33	2019	10	276	Saturday
2103	10/6/2019	4.00	11.690	2.00	34.60	6.0	0.0	5.0	4.20	2019	10	276	Sunday
2104	10/7/2019	7.34	4.507	3.00	50.80	6.0	0.0	10.0	1.00	2019	10	276	Monday
2105	10/8/2019	0.33	1.730	0.50	44.30	20.0	2.0	2.0	0.00	2019	10	190	Tuesday

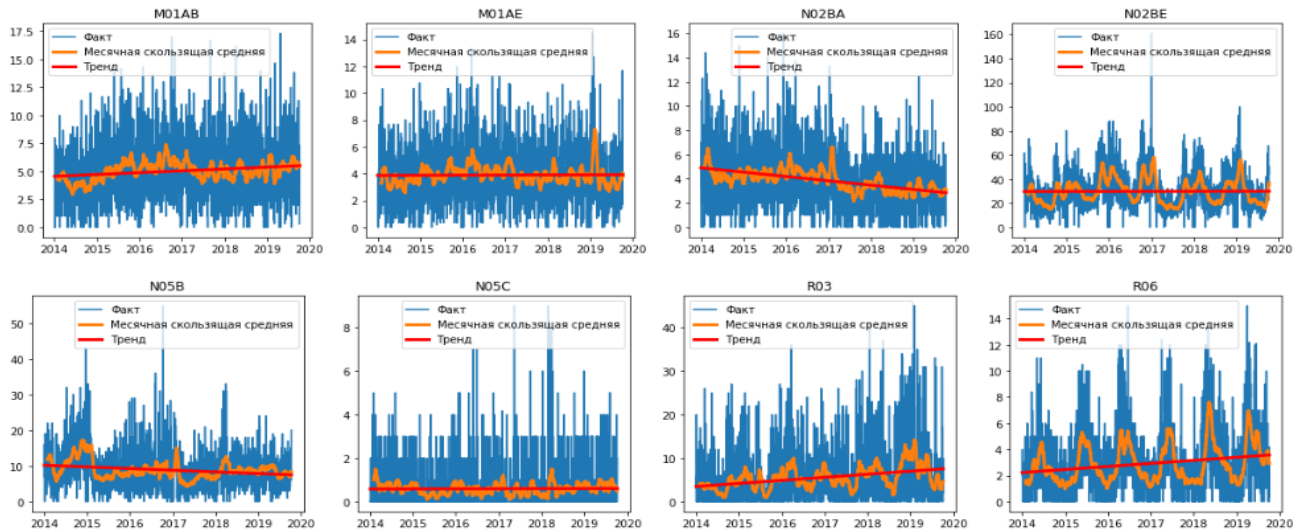
2106 rows × 13 columns

Рисунок 2. Фрагмент таблицы датасета

Решение автоматически воспроизводит аналитический материал для пользователя в виде краткой сводки о том, с какими входными данными он имеет дело. Разведочный анализ данных из текущего датасета с прилагающимися описаниями визуализаций выглядит следующим образом (*Рисунок 3, Рисунок 4, Рисунок 5, Рисунок 6*):

```
plt.subplot(1, 4, (i)%4+1)
plt.plot(dynam['datum'], dynam[view], label = 'Факт')
plt.plot(dynam['datum'], dynam[view].rolling(30).mean(), label = 'Месячная скользящая средняя', LineWidth = 3)
plt.title(view)
z = np.polyfit(range(dynam['datum'].shape[0]), dynam[view], 1)
p = np.poly1d(z)
plt.plot(dynam['datum'], p(range(dynam['datum'].shape[0])), LineWidth = 3, color = 'red', label = 'Тренд')
plt.legend()
if (i)%4 == 0:
    plt.show()
```

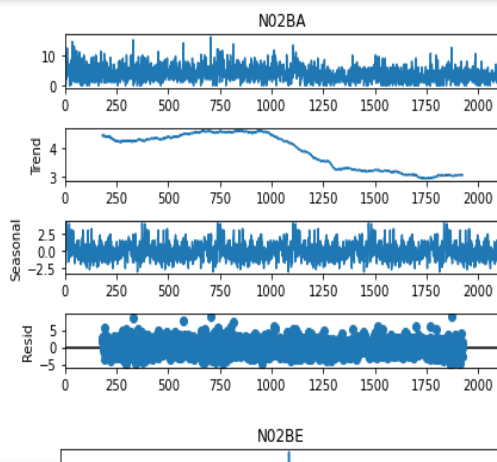
ДИНАМИКА ПРОДАЖ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:



Как можно заметить: все паттерны фактов очень разные, что говорит о том, что, скорее всего, будут подходить разные методы прогнозирования для разных категорий товаров. Есть как восходящие тренды, так и затухающие, а также и практически полное отсутствие направленного движения ряда данных вовсе. Тем не менее, у всех динамик фактов есть сезонность: как аддитивная, так и мультипликативная, что говорит о том, что стоит рассматривать только те методы прогнозирования, которые способны учитывать сезонные колебания.

Рисунок 3. Разведочный анализ (1)

```
In [254]: #Декомпозиция временных рядов категорий лекарств:
print('\033[1m' + 'ДЕКОМПОЗИЦИЯ ВРЕМЕННЫХ РЯДОВ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:' + '\033[0m')
for _, view in enumerate(col):
    result = seasonal_decompose(dynam[view], model='additive', period=365)
    result.plot()
    plt.show()
```

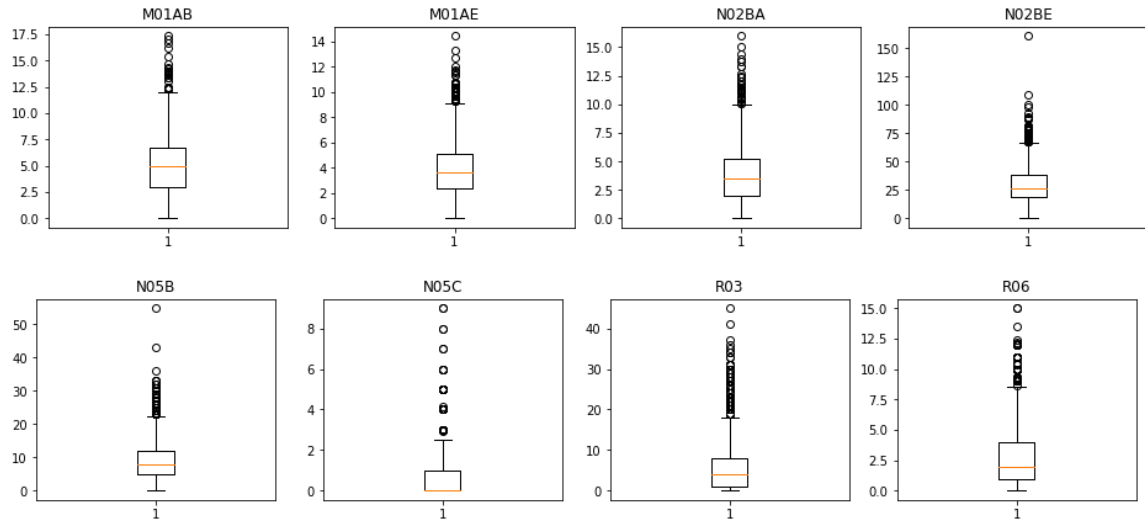


При декомпозиции временных рядов разных категорий лекарств можно прийти к выводу о том, что при разбиении на тренд-сезонная модель хорошо объясняет их поведение. Колебания остатков происходят в довольно ограниченной амплитуде практически во всех категориях, кроме "N02BE", что говорит о том, что поведение рассматриваемых временных рядов не совсем хаотично и может быть объяснено характером их компонент.

Рисунок 4. Разведочный анализ (2)

```
In [256]: #Построение "ящиков с усами" для выявления выбросов в данных:
print('\033[1m' + '"ЯЩИКИ С УСАМИ" ПРОДАЖ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:' + '\033[0m')
for _, view in enumerate(col):
    if (_)%4+2 == 2:
        plt.figure(figsize=(16,3))
        plt.subplot(1, 4, (_)%4+1)
        plt.boxplot(dynam[view])
        plt.title(view)
    if (_)%4+2 == 0:
        plt.show()
```

"ЯЩИКИ С УСАМИ" ПРОДАЖ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:

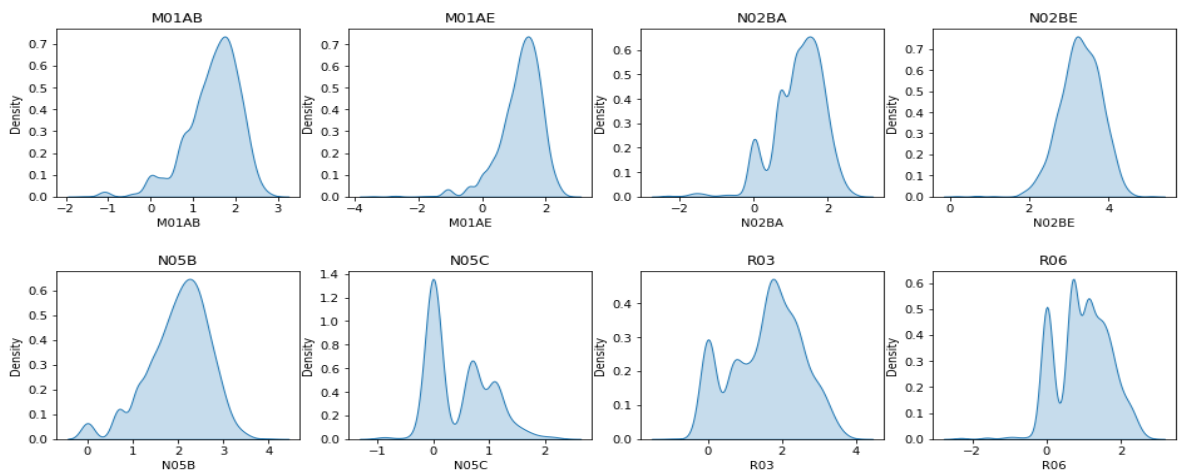


При условии нескольких тысяч точек значений выбросы единичны. Наиболее существенный выброс наблюдается в ранее уже упомянутом "проблемной" категории "N02BE". Тем не менее, датасет сформирован из реальных данных, поэтому никакие записи удалять нельзя. Можно считать, что иногда, по каким-то причинам, бывают аномальные изменения величины спроса на разные категории лекарств, и будущая прогнозная модель должна достойно работать в рамках этого допущения.

Рисунок 5. Разведочный анализ (3)

```
In [17]: #Построение гистограмм распределений:
print('\033[1m' + 'РАСПРЕДЕЛЕНИЕ ФАКТОВ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:' + '\033[0m')
for _, view in enumerate(col):
    if (_)%4+2 == 2:
        plt.figure(figsize=(16,3))
        plt.subplot(1, 4, (_)%4+1)
        sns.kdeplot(np.log(dynam[view]), shade=True)
        plt.title(view)
    if (_)%4+2 == 0:
        plt.show()
```

РАСПРЕДЕЛЕНИЕ ФАКТОВ РАЗНЫХ КАТЕГОРИЙ ЛЕКАРСТВ:



Как можно заметить: все факты категорий лекарств, кроме "N05C", "R03" и "R06", имеют распределение, близкое к нормальному, поэтому работа прогнозных моделей с ними будет упрощена. В категориях "N05C", "R03" и "R06" имеет место быть дублирование с ассиметрией, которая, скорее всего, была вызвана наличием нескольких "аномальных" значений в некоторые дни продаж. Тем не менее, ввиду реальности данных прогнозирование будет проводиться на неизменённом датасете.

Рисунок 6. Разведочный анализ (4)

2) Математический аппарат:

В рамках выполнения процедуры прогнозирования использовались 1 метод разбиения месячного прогноза на дневной с помощью ежедневных коэффициентов, 3 статистических методов прогнозирования (Метод Хольта-Уинтерса, ETS-модели, SARIMA) и 1 нейронная сеть «LSTM» (Long Short-Term Memory). Разберём механику всех вышеупомянутых инструментов по порядку:

- **Метод получения дневного прогноза:**

В основе метода лежит линейная оптимизация коэффициентов умножения в сумме предыдущего месяца предыдущего месяца первого месяца месячного прогноза и предыдущего месяца прошлого года первого месяца месячного прогноза с целью достижения минимальной разницы между получившимся новым месяцем путём ранее изложенной операции и предыдущего месяца первого месяца месячного прогноза при положительных коэффициентах. Значения подобранных коэффициентов сохраняются и используются для умножения в сумме предыдущего месяца первого месяца месячного прогноза и текущего месяца в прошлом году первого месяца месячного прогноза. После чего полученный вектор нормируется и становится набором упорядоченных по датам весов конкретного дня, которые впоследствии будут умножаться на значения первого месяца месячного прогноза с целью его разбивки на суточные разрезы. Графически данный метод можно объяснить следующим образом на примере декомпозиции августа 2019 года (*Таблица 2*):

Таблица 2. Метод получения дневного прогноза

№ шага	Действие	Примечание
1	$k1 \cdot [2019-06] + k2 \cdot [2018-07] - [2019-07]$	k1 и k2 линейно оптимизируются для минимизации разницы (k1 и k2 могут быть только положительными)
2	$k1 \cdot [2019-07] + k2 \cdot [2018-08] = [2019-08]$	применение сохранённых k1 и k2 для "прогнозирования" первого месяца месячного прогноза
3	[каждое значение] в [2019-08] / сумма([2019-08])	получение нормированного вектора с "весами" дней
4	[значение первого месяца месячного прогноза] * [вектор с "весами" дней] (из шага № 3)	распределение месячного прогноза на дни

- **Метод Хольта-Уинтерса.** Данный метод прогнозирования — это модификация метода Хольта, и он основывается на представлении динамики развития временного ряда в качестве линейной зависимости с постоянно изменяющимися параметрами, которыми являются, помимо уровня временного ряда (L) и его

тренда (Т), ещё и сезонность (S). Параметры модели прогнозирования последовательно высчитываются в течение всего временного ряда, останавливаясь на его последнем значении, после которого они становятся основой для её функционирования. Как правило, этот инструмент используется для временных рядов с трендом и сезонными колебаниями, что делает его модель прогнозирования более сложной и чувствительной к неустойчивостям. Математические модели данного метода прогнозирования зависят от типа сезонности и описываются следующими формулами:

$$\text{Аддитивная сезонность: } \check{Y}_{t+1} = L_t + T_t + S_t \quad (1.1)$$

$$L_t = \alpha * (Y_t - S_{t-m}) + (1 - \alpha) * (L_{t-1} + T_{t-1}), \quad (1.1.1)$$

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1}, \quad (1.1.2)$$

$$S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-m}, \quad (1.1.3)$$

$$\text{Мультипликативная сезонность: } \check{Y}_{t+1} = (L_t + T_t) * S_t \quad (1.2.1)$$

$$L_t = \alpha * (Y_t / S_{t-m}) + (1 - \alpha) * (L_{t-1} + T_{t-1}), \quad (1.2.2)$$

$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1}, \quad (1.2.3)$$

$$S_t = \gamma * (Y_t / L_t) + (1 - \gamma) * S_{t-m}, \quad (1.2.4)$$

\check{Y}_{t+1} – прогноз на следующий период $t+1$;

Y_t – фактическое значение временного ряда в момент времени t ;

m – длина сезонного цикла (день, неделя, месяц, квартал, год и т.д.),
 $m < t$;

L_t – сглаженное значение временного ряда в момент времени t ;

α – коэффициент сглаживания временного ряда ($0 \leq \alpha \leq 1$);

T_t – сглаженное значение линейного тренда в момент времени t ;

β – коэффициент сглаживания линейного тренда ($0 \leq \beta \leq 1$);

S_t – сезонный коэффициент, присущий моменту времени t (для аддитивного типа сезонности: $S_{t \in [1; m]} = Y_{t \in [1; m]} - L_1$, а $S_{t \in (m; +\infty)}$; для мультипликативного типа сезонности $S_{t \in [1; m]} = Y_{t \in [1; m]} / L_1$, а $S_{t \in (m; +\infty)}$);

γ – коэффициент сглаживания сезонных колебаний ($0 \leq \gamma \leq 1$).

- **ETS-модели (error-trend-seasonality).** Данный метод прогнозирования основывается на принципе экспоненциального сглаживания. Если в ранее рассмотренном простом экспоненциальном сглаживании процедуре сглаживания подвергался только уровень временного ряда, то данные модели способны проводить ту же самую операцию и с трендом, и с сезонностью. Механизм сглаживания определяется совокупностью типов трёх компонент временного ряда, к которому применяется указанный инструмент прогнозирования. Таким образом, общее обозначение метода ETS с параметрами-предикторами приобретёт следующий вид:

$$ETS_{(\text{ошибка, тренд, сезонность})} \quad (2)$$

ошибка (остатки): аддитивные (A), мультипликативные (M);

тренд: аддитивный (A), мультипликативный (M), аддитивный убывающий (Ad), мультипликативный убывающий (Md), отсутствующий (N);

сезонность: аддитивный (A), мультипликативная (M), отсутствующая (N).

- **SARIMA** . Данный метод прогнозирования является производным от ARIMA и является её сезонной версией: приставка «S-» означает «Seasonal». Эта модифицированная модель формируется с помощью дополнительных предикторов-параметров, которые дополнительно добавляются к базовой модели:

p: порядок авторегрессионной части;

d: порядок разностей, необходимый для приведения временного ряда в стационарный вид, если это необходимо (если временной ряд является стационарным, то его автокорреляция в модели зависит только от разности временных интервалов, а не от самого временного интервала);

q: порядок скользящего среднего;

P: сезонный порядок авторегрессии;

D: порядок сезонных разниц;

Q: сезонный порядок скользящих средних;

m: количество временных циклов в сезонном периоде.

Таким образом, общее обозначение метода SARIMA приобретёт следующий вид:

$$SARIMA_{(p,d,q) (P,D,Q)m} \quad (3)$$

- **Нейронная сеть «LSTM» [4].** Данный тип нейронной сети является модификацией рекуррентных нейронных сетей, которые способны передавать

информацию своим копиям. Эта отличительная черта позволяет ей успешно работать с временными рядами, помогая «запоминать» свои прошлые предсказания и их метрики, а собственно, и продолжать экстраполяцию на будущие периоды более успешно. Принцип её работы хорошо продемонстрирован с помощью данного gif-файла, в котором показано, как сеть сохраняет прошлые результаты (*Рисунок 7*) [5]:

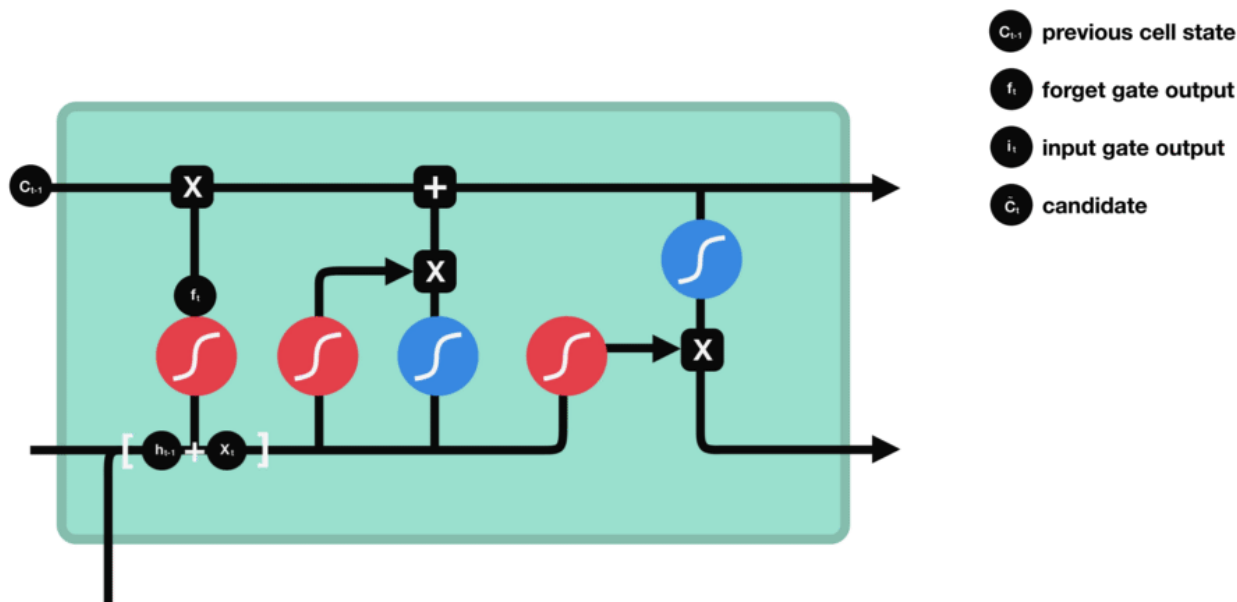


Рисунок 7. Принцип работы LSTM-сети

Стоит отметить, что подбор параметров для статических методов прогнозирования ограничился только перебором аддитивной и мультипликативной сезонностей на тестовых периодах с целью выявления лучших значений ошибок прогнозирования и сохранения типа указанной сезонности для прогнозирования полного периода временного ряда. Тем не менее, подбор параметров для нейронной сети осуществлялся более экспериментальным образом во время попыток уменьшения значений величин ошибок прогнозирования на тестовых периодах. Для нейронной сети, LSTM, которая используется как отдельный метод прогнозирования в рассматриваемом решении, список параметров может быть выражен следующим словарём (*Рисунок 8*):

```
params = {  
    "loss": "mean_squared_error",  
    "optimizer": "adam",  
    "dropout": 0.1,  
    "lstm_units": 300,  
    "epochs": 50,  
    "batch_size": 2,  
    "es_patience": 10  
}
```

Рисунок 8. Словарь параметров LSTM

3) Выходные данные:

После работы цикла со всей обширной процедурой прогнозирования, основанной на временном окне, указанном пользователем, и входящем датасете с временными рядами у пользователя появляется возможность увидеть результаты сессии данного ИТ-решения. Выходными объектами являются таблица со статистическими ошибками методов прогнозирования на тестовых периодах (*Рисунок 9*), таблица распределения временных рядов с их оптимальными методами прогнозирования (*Рисунок 10*), набор графиков с фактами временных рядов и прогнозами, построенными с помощью оптимальных для них моделей прогнозирования (*Рисунок 11*), месячный прогнозный план (*Рисунок 12*) и дневная декомпозиция первого месяца месячного прогнозного плана (*Рисунок 13*). Данные наглядные материалы позволяют пользователю принимать управленческие решения в рамках предпринимательской деятельности, а также сделать запрос на изменение механики работы процедуры в случае обнаружения каких-либо выгодных закономерностей. Также выводится значение полного времени работы ИТ-решения (*Рисунок 14*), что позволяет оценить загруженность ресурсов при решении каких-либо задач.

```
#Вывод таблицы со статистическими ошибками:
pd.set_option('display.max_columns', None)
print('\033[1m' + 'СТАТИСТИЧЕСКИЕ ОШИБКИ НА ТЕСТОВЫХ ПЕРИОДАХ:' + '\033[0m')
errors_cul
```

СТАТИСТИЧЕСКИЕ ОШИБКИ НА ТЕСТОВЫХ ПЕРИОДАХ:

	Метод	MAE (M01AB)	RMSE (M01AB)	MAPE,% (M01AB)	MAE (M01AE)	RMSE (M01AE)	MAPE,% (M01AE)	MAE (N02BA)	RMSE (N02BA)	MAPE,% (N02BA)	MAE (N02BE)	RMSE (N02BE)	MAPE,% (N02BE)	MAE (N05B)	RMSE (N05B)	MAPE,% (N05B)
0	Хольт-Уинтерс	13.78	16.46	8.92	13.87	21.58	10.17	9.25	12.27	10.38	108.00	161.91	10.55	43.73	48.20	16.91
1	ETS	13.78	16.46	8.92	13.87	21.58	10.17	9.25	12.27	10.38	118.30	166.95	12.15	43.73	48.23	16.91
2	SARIMA	21.78	26.90	13.12	14.77	26.17	10.10	19.94	24.08	20.38	477.95	555.65	48.18	72.49	80.08	26.77
3	LSTM	11.73	15.33	7.20	17.18	23.99	13.37	32.54	33.95	36.54	563.87	635.07	57.57	27.14	33.93	10.19
4	Комбинация 2-х лучших методов	12.44	14.97	7.85	13.87	21.58	10.17	9.25	12.27	10.38	113.00	163.78	11.33	24.39	30.69	9.49

Рисунок 9. Выходные данные (1)

```
#Вывод таблицы с выбранными моделями прогнозирования для разных категорий:
print('\033[1m' + 'ОПТИМАЛЬНЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ДЛЯ КАЖДОЙ КАТЕГОРИИ:' + '\033[0m')
tools_cul.columns = ['Категория', 'Оптимальная модель']
tools_cul = tools_cul.reset_index(drop = True)
tools_cul
```

ОПТИМАЛЬНЫЕ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ДЛЯ КАЖДОЙ КАТЕГОРИИ:

	Категория	Оптимальная модель
0	M01AB	LSTM
1	M01AE	SARIMA
2	N02BA	Хольт-Уинтерс
3	N02BE	Хольт-Уинтерс
4	N05B	Комбинация Хольта-Уинтерса и LSTM
5	N05C	LSTM
6	R03	ETS-модель
7	R06	Хольт-Уинтерс

Рисунок 10. Выходные данные (2)

```
col.pop(0)

for _, view in enumerate(col):
    if (_)%4+2 == 2:
        plt.figure(figsize=(20, 4))
        plt.subplot(1, 4, (_)%4+1)
        plt.plot(dynam_full['datum'], dynam_full[view], label = 'Факт', linewidth = 5)
        plt.plot(pd.to_datetime(long_cul_full['datum'], errors='coerce'), long_cul_full[view], label = 'Прогноз', linewidth = 5)
        plt.legend()
        plt.title(view)
        if (_)%4+2 == 0:
            plt.show()
```

ВИЗУАЛИЗАЦИЯ ПРОГНОЗОВ ОПТИМАЛЬНЫХ МОДЕЛЕЙ ДЛЯ КАЖДОЙ КАТЕГОРИИ:

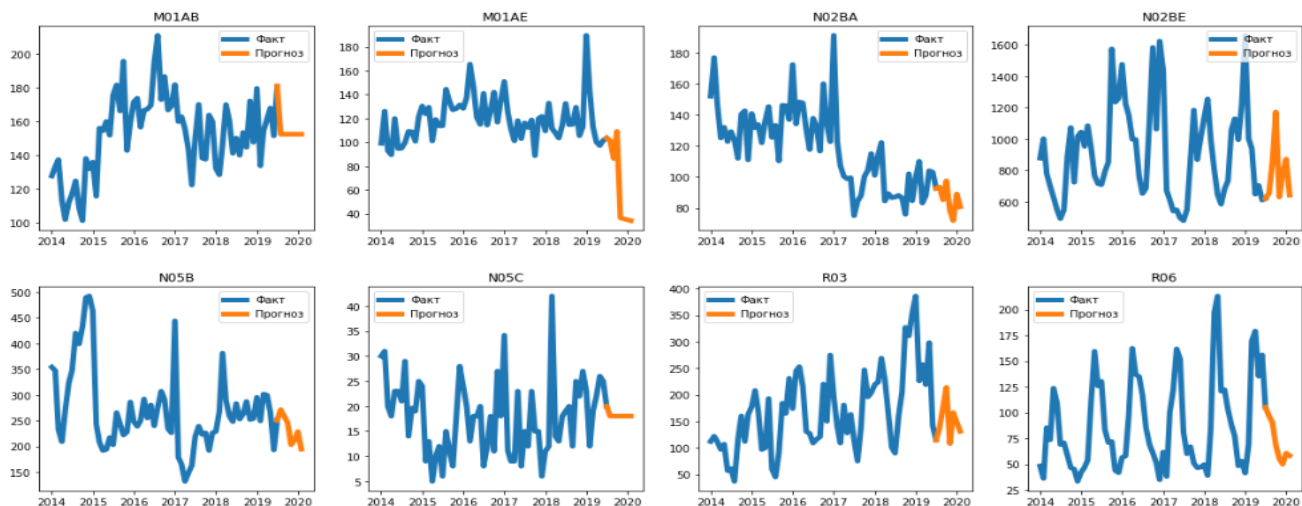


Рисунок 11. Выходные данные (3)

```
#Вывод таблицы с прогнозным планом:
pd.set_option('display.max_rows', None)
long_cul.rename(columns={'datum': 'Дата'}, inplace=True)

print('\033[1m' + 'ПРОГНОЗНЫЙ ПЛАН:' + '\033[0m')
long_cul[col] = long_cul[col].astype('float64').round(2)
long_cul['Дата'] = long_cul['Дата'].dt.strftime('%Y-%m')
long_cul
```

ПРОГНОЗНЫЙ ПЛАН:

	Дата	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
0	2019-08	152.51	101.37	93.67	653.71	271.28	18.02	137.49	96.86
1	2019-09	152.51	86.42	85.35	853.80	258.44	18.02	176.51	90.37
2	2019-10	152.51	109.10	97.63	1172.48	246.42	18.02	214.08	68.53
3	2019-11	152.51	36.52	78.60	630.18	203.18	18.02	108.13	55.51
4	2019-12	152.51	35.75	71.99	780.22	211.89	18.02	166.45	50.02
5	2020-01	152.51	35.00	89.17	871.62	228.47	18.02	148.34	60.66
6	2020-02	152.51	34.26	80.91	642.08	194.53	18.02	131.18	58.39

Рисунок 12. Выходные данные (4)

```
print('\033[1m' + 'ДЕТАЛИЗАЦИЯ ПЕРВОГО МЕСЯЦА ПРОГНОЗНОГО ПЛАНА:' + '\033[0m')
short_cul
```

ДЕТАЛИЗАЦИЯ ПЕРВОГО МЕСЯЦА ПРОГНОЗНОГО ПЛАНА:

	Дата	M01AB	M01AE	N02BA	N02BE	N05B	N05C	R03	R06
0	2019-08-01	2.39	2.01	1.04	13.50	8.74	1.02	7.82	5.56
1	2019-08-02	1.82	2.00	1.70	12.04	10.89	0.00	1.30	4.44
2	2019-08-03	3.99	2.59	2.93	24.57	9.69	0.00	0.96	1.90
3	2019-08-04	5.08	2.46	1.23	10.10	9.94	2.94	2.87	6.98
4	2019-08-05	5.43	2.03	4.77	15.88	7.64	2.03	1.04	1.44
5	2019-08-06	7.01	3.40	2.50	18.06	5.15	0.00	2.00	2.19
6	2019-08-07	3.39	4.50	6.19	19.24	5.15	1.02	3.04	3.45
7	2019-08-08	4.85	3.45	2.08	15.08	12.19	-0.12	3.82	2.51
8	2019-08-09	2.66	2.68	4.02	21.48	13.52	-0.47	2.87	1.87
9	2019-08-10	7.26	4.62	4.58	18.61	7.26	0.00	0.96	2.51
10	2019-08-11	3.77	4.18	5.39	27.09	12.94	1.02	0.87	7.06
11	2019-08-12	10.94	4.72	5.41	24.02	7.03	0.00	2.08	1.87
12	2019-08-13	3.85	3.85	1.65	19.73	11.38	-0.35	0.52	2.84
13	2019-08-14	2.83	3.94	4.77	16.82	6.83	0.00	5.47	3.37
14	2019-08-15	7.98	2.59	3.16	27.14	10.66	1.92	0.00	2.87
15	2019-08-16	4.49	5.31	3.73	20.03	6.94	3.05	2.78	4.34
16	2019-08-17	2.85	1.30	2.54	23.61	6.34	0.90	0.87	1.26
17	2019-08-18	7.80	3.93	0.55	27.09	8.24	1.02	3.91	0.97
18	2019-08-19	6.62	4.65	2.32	23.06	8.95	0.00	3.73	5.92
19	2019-08-20	4.09	4.66	3.00	18.63	6.11	0.00	9.21	2.48
20	2019-08-21	3.62	1.41	1.79	20.40	4.78	1.02	1.13	3.41
21	2019-08-22	3.69	3.17	0.85	15.57	7.17	1.02	0.00	4.38
22	2019-08-23	3.80	2.01	4.67	20.51	10.04	0.00	10.68	2.60
23	2019-08-24	5.27	2.94	1.47	33.49	13.26	0.00	5.56	2.76
24	2019-08-25	4.40	5.33	4.09	22.70	6.69	-0.12	11.55	2.17
25	2019-08-26	11.65	4.83	2.50	35.89	11.72	0.00	9.03	3.59
26	2019-08-27	3.32	0.41	5.95	18.29	6.46	-0.35	2.08	1.71
27	2019-08-28	2.74	1.80	2.32	19.07	5.63	0.90	0.00	2.02
28	2019-08-29	7.39	3.37	1.65	26.03	11.72	1.02	1.82	2.48
29	2019-08-30	4.62	4.38	3.97	27.58	8.49	-0.47	32.22	2.84
30	2019-08-31	2.92	2.86	0.85	18.40	9.71	1.02	7.30	3.08

Рисунок 13. Выходные данные (5)

```
end_time = datetime.now()
print('\033[1m' + 'ВРЕМЯ ВЫПОЛНЕНИЯ ИТ-РЕШЕНИЯ: ' + '\033[0m')
print('Duration: {}'.format(end_time - start_time))
```

ВРЕМЯ ВЫПОЛНЕНИЯ ИТ-РЕШЕНИЯ:
Duration: 0:13:07.837175

Рисунок 14. Выходные данные (6)

Таким образом, выходные данные позволяют нам дать оценку успешности выполнения алгоритма и измерить эффективность его функционирования. Более того, они отлично играют свою главную роль – поддержку при принятии управленческих решений, касаемых вопросов управления ресурсами и планирования деятельности предприятия.

Заключение

В заключение хотелось бы сказать, что цель проекта можно считать достигнутой: удалось эффективно использовать методы статистики и машинного обучения для получения точных прогнозных планов в разных временных разрезах на основе процедуры прогнозирования временных рядов. Алгоритм обладает такими свойствами, как итеративность и универсальность, поэтому может работать с любыми данными подобного типа. В рамках выполнения проекта удалось познакомиться с теоретическими аспектами, относящимися к проведению процедуры прогнозирования временных рядов, а также получить практический опыт в реализации всего теоретического потенциала в создании конкретного решения, оперирующего на его постулатах.

Также хотелось бы отметить, что текущая версия алгоритма может быть использована для следующий наработок. Например, возможно добавление новых моделей прогнозирования или формирование новых выходных таблиц с необходимым аналитическим материалом. В связи с этим можно заявить, что дальнейшие работы в данной сфере будут только продолжаться.

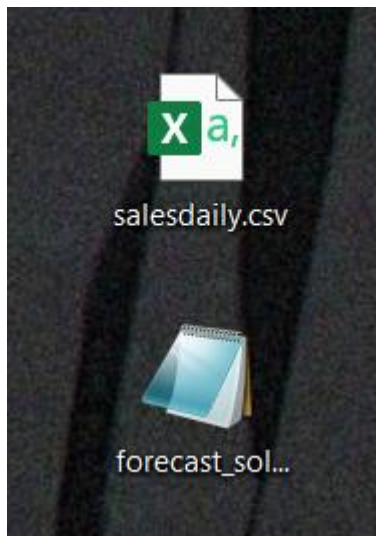
Список используемых источников

- 1) Makridakis S, Spiliotis E, Assimakopoulos (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 13(3): e0194889. Н, 26 p.;
- 2) Aiolfi M., Timmermann A. (2006). Persistence in forecasting performance and conditional combination strategies. Journal of Econometrics 135, 31-53 p.;

- 3) Социальная сеть специалистов по обработке данных и машинному обучению «Kaggle» [Электронный ресурс]//Режим доступа:
<https://www.kaggle.com/datasets/milanzdravkovic/pharma-sales-data>;
- 4) Информационный портал про машинное обучение и программирование в целом «GeeksForGeeks» [Электронный ресурс]//Режим доступа:
<https://www.geeksforgeeks.org/long-short-term-memory-networks-explanation>;
- 5) Отечественный информационный портал про машинное обучение и программирование в целом «IT CHEF.RU» [Электронный ресурс]//Режим доступа: <https://itchef.ru/articles/30183/>;

Примечания

- Набор файлов:



- Код алгоритма находится в файле с расширением «.ipynb» в том же архиве, в котором находится отчёт.

