

УДК 004.622

Разработка алгоритмов предобработки информации для прогнозных моделей ИСППР управления закупками

Н.К. Андриевская, Т.В. Мартыненко
Донецкий национальный технический университет
nataandr@yandex.ua

Аннотация

Статья посвящена разработке алгоритмов предварительной обработки данных и решению таких задач, как устранение неполноты данных и преобразование необработанной информации в удобный для анализа вид. При получении реальных данных временных рядов мы сталкиваемся с тем, что они лишь изредка бывают в необходимом формате. Существенная часть необработанных наборов данных слабо структурирована или содержит большое количество отсутствующих значений. Это делает невозможным последующую обработку данных некоторыми методами. В связи с этим реализованы основные этапы предварительной обработки данных, такие как, очистка данных, заполнение недостающих значений, сглаживание временного ряда, нормализация переменных. Таким образом, получен набор данных для дальнейших исследований, реализации и тестирования прогнозных моделей интеллектуальной системы поддержки принятия решений (ИСППР) управления закупками.

Общая постановка проблемы

В современном мире, в условиях быстро изменяющихся рынков, невозможно достичь ощутимых конкурентных преимуществ без использования современных информационных технологий. Наиболее передовыми в отношении использования современных информационных систем являются предприятия сферы бизнеса, в первую очередь крупные торговые структуры, в том числе и аптечные холдинги.

Фармацевтический сектор экономики является важным элементом, как в социальном, так и в экономическом плане. Существенными особенностями реализации лекарственных средств является многообразие ассортимента и высокая динамика его изменения, как в номенклатуре, так и в объемах и ценах. Все это влечет за собой требования в постоянном мониторинге реализации лекарственных средств, соответствующем анализе, использовании различных методов моделирования различных ситуаций и прогнозирования, поддержки принятия определенных решений в объемах в ценовой политике по закупке лекарственных средств аптечными холдингами и отдельными аптеками [1].

Важнейшей инновацией в этой области является способность ключевых бизнес-процессов к цифровизации, которая позволяет ускорить выполнение бизнес-процессов в закупочных процессах, обеспечивая большую достоверность и прозрачность информации для принятия обоснованных решений [2].

Закупочная деятельность - одна из основных коммерческих функций любого торгового предприятия, в том числе и аптечной сети. Правильно организованная и четко выстроенная закупочная стратегия не только способствует удовлетворению потребительского спроса, но и позволяет уменьшить риск, связанный с плохим сбытом товаров.

В настоящее время организация процесса управления запасами имеет огромное значение в фармацевтической отрасли, так как она направлена на поддержание необходимого и достаточного объема запасов для выполнения объема продаж, а также прямо влияет на безопасность людей [3].

Анализ и принятие решений менеджерами привычными методами затрудняется из-за огромного объема информации и большого количества характеристик для анализа. Принятие решений становится чрезвычайно ресурс-затратным процессом, снижается точность, повышается время, необходимое для решения основных задач, в том числе и по планированию закупок. Среди популярных рыночных предложений в настоящее время на рынке программного обеспечения для аптек представлено несколько видов продуктов: «Аналит-аптека» («Аналит»), «ФармаНет» («Фарманет»), СПО «еФарма», «АИС Аптекарь», «АИС Парацельс», «АИС 1С: Аптека».

Поскольку ни одно из выше описанных программных решений не обеспечивает требования заказчика по функциональности СППР, необходимой и актуальной становится

задача разработки интеллектуальной системы поддержки принятия решений для закупочной деятельности аптеки, построенной на применении целого ряда информационных технологий.

Постановка задачи исследований

В более раннем исследовании по этой теме при помощи имитационных моделей было выполнено моделирование в аспекте динамики изменения финансовых показателей аптеки в зависимости от различных стратегий организации закупок и выбрана оптимальная стратегия, которая и была реализована в информационной системе [4].

На сегодняшний день одним из ведущих направлений в сфере информационных технологий являются интеллектуальные информационные системы. Эти системы позволяют решать неформализованные задачи, возникающие в ходе развития бизнеса.

Закупочная функция также эволюционирует под воздействием развития цифровых технологий, что означает внедрение принципиально новых средств анализа и обработки данных в систему управления предприятием.

Постоянной функцией менеджера по закупкам аптеки является проведение анализа закупочных цен поставщиков и определение «оптимальной» цены, наиболее подходящей в данный момент для покупки товара. Возникает необходимость прогнозирования данных (цен) по историческому массиву цен из прайс-листов.

Реализация поставленной выше задачи предусматривает использование целого комплекса информационных технологий, применения различных математических методов и моделей, в том числе статистических и интеллектуальных. При этом возможно решение одной и той же задачи различными методами, что усложняет реализацию системы.

В работе [5] описано прогнозирование закупочных цен лекарств в условиях аптечной сети экстраполяционными методами прогнозирования временных рядов по типу традиционного статического подхода, при этом эксперименты показали достаточно высокую точность прогнозирования.

Представителями моделей искусственного интеллекта являются прогнозные модели на основе нейронных сетей. Использование данных моделей позволяет ассоциировать исследуемый показатель (цена на лекарственное средство) с набором различных факторов (не только исторических значений) и экстраполировать его значение на перспективу [6]. При этом задача прогнозирования закупочной цены, спроса и продаж лекарств обладает теми особенностями, которые делают целесообразным использование

именно интеллектуальных нейросетевых методов моделирования и, в частности, с топологией «внутренний учитель»: небольшой объём данных; присутствуют пропуски данных; возможны искажения данных («шум»); необходима адаптация модели при поступлении новых данных; наличие большого количества позиций номенклатуры товаров [7,8].

Основная алгоритмическая часть ИС построена на прогнозировании «допустимой» закупочной цены среди цен разных производителей и поставщиков искусственной нейронной сетью, которую по имеющимся результатам оценки эффективности закупок станет возможно переобучать, чтобы картина соответствовала действительности в большей степени.

Искусственная нейронная сеть в процессе своей работы будет делать краткосрочный дневной либо недельный прогноз цен для формирования плана закупок товаров, на основании которых одна из подсистем должна проанализировать остатки на складе и, в случае возможной нехватки, вывести информацию, каких и сколько товаров необходимо закупить.

Анализ литературных источников показал, что алгоритмам и методам предобработки данных, а также формированию датасетов, уделено чрезвычайно мало внимания разработчиками нейронных сетей. Модели и алгоритмы машинного обучения учеными-практиками описаны с точки зрения их применения на чистых, уже подготовленных данных [9].

Предварительная обработка данных – это процесс анализа данных, который включает преобразование необработанной информации в удобный для анализа формат и дальнейшее их использование. Задачи предварительной подготовки данных в первую очередь связаны с обработкой прайс-листов цен лекарств.

В результате исследований необходимо решить следующие проблемы: обеспечить одинаковый интервал между рядами наблюдений; обеспечить количество пробелов данных на уровне не более 20%; выполнить заполнение промежутков данными; устранить пиковые мгновенные значения; устранить «шум».

Обзор основных методов и технологий

Временной ряд – последовательность значений, которые протекают и измеряются в обозначенном промежутке времени. Временной ряд отличается от обычной выборки данных тем, что он учитывает не только значения, но и время, в течение которого они были измерены [10].

При анализе, в общем случае, временной ряд предусматривает наличие четырех составляющих: Тренд (Tt), Сезонность (St),

Цикличность (Ct), Случайные возмущения (Et).

Тренд — это общая тенденция, сезонность, как следует из названия - влияния периодичности (день недели, время года и т.д.) и, наконец, шум — это случайные факторы. Циклическая компонента, по сравнению с сезонностью, имеет более длительный эффект и меняется от цикла к циклу. Поэтому, ее обычно объединяют с трендом. Обычно, выделяют три модели временного ряда:

1. Аддитивная: $Y[t]=T[t]+S[t]+E[t]$
2. Мультипликативная: $Y[t]=T[t]*S[t]*E[t]$
3. Смешанная: $Y[t]=T[t]*S[t]+E[t]$

Построение этих моделей сводится к расчету тренда (T[t]), сезонности (S[t]) и случайных возмущений (E[t]) для каждого уровня ряда (Y[t]).

При выборе необходимой модели временного ряда смотрят на амплитуду колебаний сезонной составляющей. Если ее колебания относительно постоянны, то выбирают аддитивную модель. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Реальные данные часто являются непоследовательными, неполными и возможно недостающими в определенном поведении или тенденциях и могут содержать большое количество ошибок. Предварительная обработка данных – это проверенный метод решения проблем, связанных с обработкой необработанных данных. Она подготавливает данные для последующей обработки.

Преимущества предварительной обработки данных:

– Точность. При устранении отсутствующих и несогласованных значений данных, возникших по вине человека или компьютера, повышается точность вашего набора данных.

– Консистенция. В процессе предварительной обработки могут возникать дубликаты данных, которые необходимо исключить для обеспечения более согласованного анализа данных и получения надёжных результатов без искажений.

– Полнота данных. Предварительная обработка данных позволяет добавлять необходимые отсутствующие данные.

– Предварительная обработка данных обычно облегчает чтение, применение и интерпретацию информации, особенно когда используется автоматизированное программное обеспечение для машинного обучения [11].

Приведем процесс предварительной обработки данных при прогнозировании временного ряда нейросетевыми методами в виде схемы (рис. 1).



Рисунок 1 – Процесс предварительной обработки данных временного ряда

Целью первого шага является непосредственно получение данных, а также оценка полноты данных. Для этого надо оценить, достаточно ли количество имеющихся данных для решения задачи, насколько полно имеющиеся переменные описывают исследуемый процесс, и какие внешние факторы могут оказывать влияние на исследуемый процесс.

Сколько исторических данных следует использовать для тренировки модели? Необходимо начать с временных рядов, которые в несколько раз больше, чем временные интервалы прогноза, а затем проверить точность получаемого прогноза. Далее следует изменять длину истории и проводить тест повторно, чтобы определить необходимое количество данных.

На втором шаге желательно максимальным образом обеспечить полноту, корректность, непротиворечивость данных, структурированность, согласованность данных.

Для этого необходимо использовать процедуры заполнения пропусков или восстановления данных, поиск невозможных значений и дубликатов, исправление форматов, сглаживание выбросов и устранение «шума».

Методов обработки пропусков числовых данных предостаточно, например, заполнение наиболее часто встречающимся значением или предопределенной константой, а также заполнение случайными значениями из выборки.

Интерполяция – это метод, который позволяет прогнозировать неизвестные значения между двумя известными значениями. Она

может быть использована для заполнения пропусков в таблице данных, используя уже известные значения [12]. Методы сглаживания необходимы для удаления «шума» из временного ряда.

На третьем шаге выполняется приведение типов данных и кодирование номинативных переменных, нормализация данных, стандартизация данных, обогащение данных, оптимизация пространства признаков.

Задача процесса нормализации - улучшить качество работы алгоритмов за счёт приведения данных к нужному диапазону с помощью следующих методов: нормализация на максимум; нормализация на интервал; ранговая нормализация.

Задача стандартизации заключается в улучшении качества работы алгоритмов за счёт приведения данных к стандартному нормальному распределению.

Нормализация и стандартизация существенно повышают эффективность метрических алгоритмов классификации: метода ближайшего соседа (k- Nearest Neighbors), метода k-средних (k-means), метода машин опорных векторов (Support Vector Machine).

После завершения всех алгоритмов предварительной обработки данных выполняется векторизация текстов, так как алгоритмы машинного обучения предназначены для работы с числовыми данными, и необходимо выполнить преобразование текста отзыва в числовой вектор признаков.

Весь процесс обработки данных, реализации и тестирования методов прогнозирования осуществлён в среде Google Colaboratory [13].

Для разработки методов прогнозирования и предварительной обработки данных были использованы следующие библиотеки языка Python:

– NumPy – библиотека Python, применяемая для математических вычислений: от базовых функций и до линейной алгебры [14].

– Pandas – библиотека Python, предназначенная для обработки и анализа структурированных данных. Такими данными называют информацию, которая получена в результате исследований и структурированную в виде таблиц. Для работы с такими массивами данных и создан Pandas [15].

– Matplotlib – библиотека Python, предназначенная для визуализации данных. В ней можно построить двумерные (плоские) и трехмерные графики [16].

– Scikit-learn (sklearn) – это один из наиболее часто используемых пакетов Python для Data Science и Machine Learning. Он содержит функции и алгоритмы для машинного обучения: классификации, прогнозирования или разбивки данных на группы [17].

– Keras – библиотека Python, применяемая для глубокого машинного обучения. Она даёт возможность быстрее создавать и настраивать модели – схемы, по которым распространяется и подсчитывается информация при обучении [18].

Результаты исследований

Данные, которые будут использоваться для прогнозирования цен (аналогично также продаж) взяты с источника Интернет и представляют собой датасет цен некоторой российской компании с 22.11.2011 по 30.12.23 в размере 2778 строк. Было принято решение взять его для обучения и тестирования рассмотренных методов и моделей, так как самым главным критерием является проверка работоспособности и точности разработанных алгоритмов. Отобрав данные, получаем в итоге набор данных со значениями цен и объёмов продаж одного лекарства. В результате подготовки набора данных для прогнозирования и моделирования, мы получили временной ряд в интервале от 22.11.2011 по 30.12.2022 со значениями цен и объёмов продаж по одному товару.

А. Разложение ряда по компонентам

Временной ряд подготовленных данных за последний год можно разделить на четыре компонента, представленные на рис. 2:

– уровень (**level**)– это среднее значение ряда;

– тренд (**trend**) – показывает, увеличивается или уменьшается значение со временем;

– сезонность (**seasonal**) – это краткосрочный повторяющийся цикл;

– шум (**resid**)– это случайное изменение, несвязанное с другими компонентами ряда [19].

Все эти составляющие можно разложить с помощью функции `seasonal_decompose` из модуля `statsmodels` в Python.

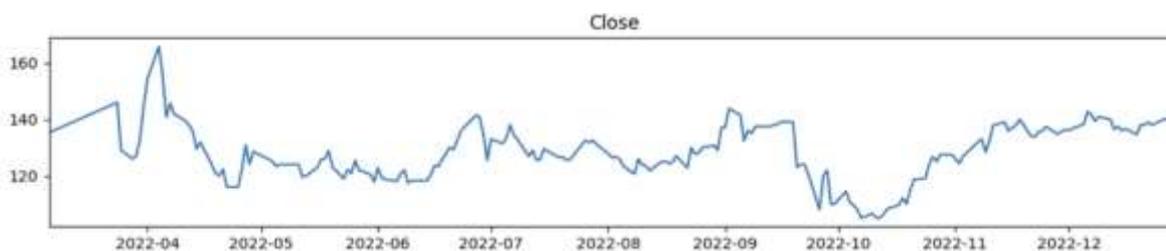


Рисунок 2 – Результат работы функции `seasonal_decompose`

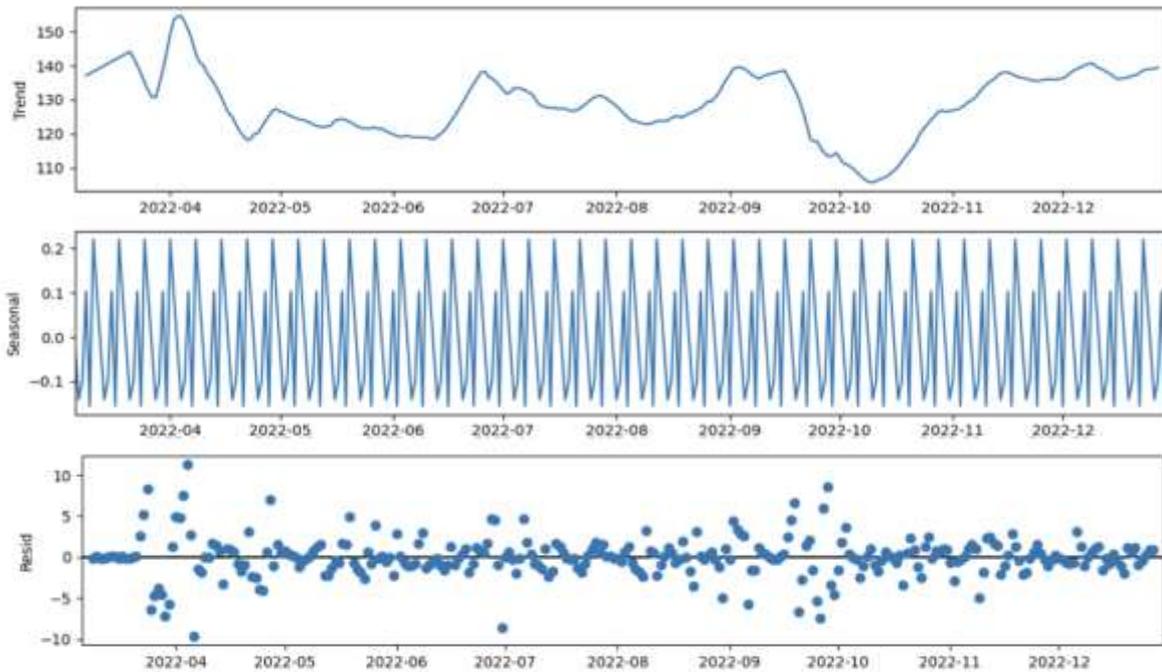


Рисунок 2 – Результат работы функции `seasonal_decompose` (продолжение)

В. Очистка данных

При получении реальных данных временных рядов мы сталкиваемся с тем, что они лишь изредка бывают в необходимом формате. Большая часть необработанных наборов данных слабо структурирована или содержит большое количество отсутствующих значений, что делает невозможным последующее обучение моделей.

Над набором данных была выполнена чистка значений дат и цен/объемов продаж. Числа приведены к одинаковому виду, все даты приведены к одному формату с помощью функции `to_datetime` библиотеки `Pandas` в `Python`.

Пример:

```
dff['Date'] =
pd.to_datetime(dff.Date,format='%Y-%m-%d')
```

С. Заполнение недостающих значений

Проанализировав данные, можно заметить, что в выборке присутствуют некоторые пропуски в датах по каким-либо причинам (множество пропущенных дат являются выходными днями или праздниками). Для более точного понимания тренда временного ряда и достижения хорошего результата прогнозирования необходимо заполнить пропущенные даты и значения для них.

Процесс заполнения пропущенных дат и достигнутый результат приведен на рисунке 3.

Таким образом, недостающие значения цен/объемов продаж были заполнены с помощью функции `interpolate` библиотеки `pandas` в `Python`.

В результате преобразования размер выборки после заполнения увеличился до 4058 записей.

Очищенный график представлен на рис. 4.



Рисунок 3 – Результат заполнения пропусков

Д. Сглаживание временного ряда

При работе с рядами шум может затруднять анализ структуры ряда, соответственно можно использовать методы сглаживания ряда для его исключения, что позволит лучше рассмотреть структуру ряда.

Сглаживание ряда – это представление тренда в определенной точке с помощью усредненного значения ряда, которое рассчитывается в окрестностях этой точки [20]. Результат сглаживания представлен на рис. 5.

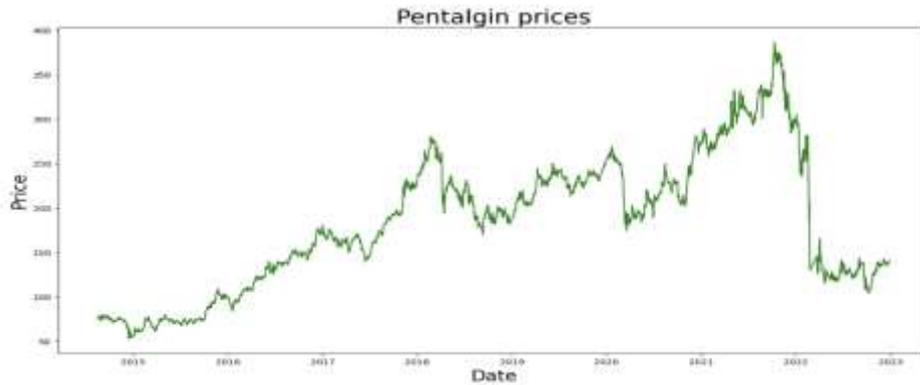


Рисунок 4 – Очищенный график выборки данных

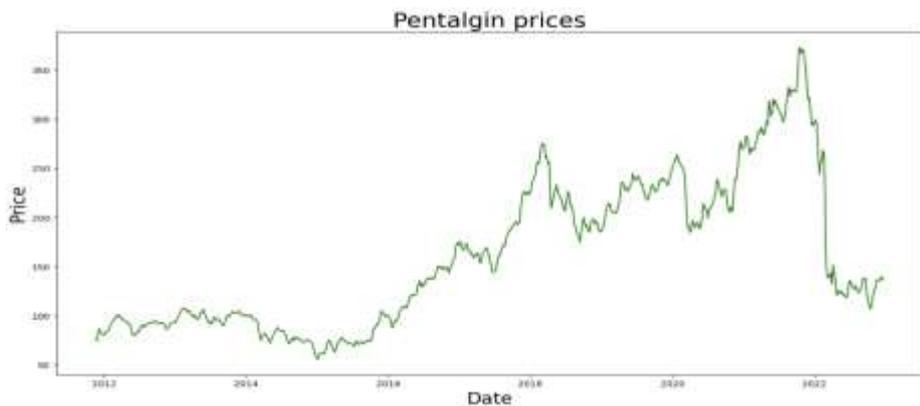


Рисунок 5 – Сглаженный график выборки данных

Для сглаживания ряда использован один из выше представленных методов прогнозирования – метод экспоненциального сглаживания. Коэффициент сглаживания выбран равным 0,2.

F. Нормализация переменных

Нормализация – это один из этапов предварительной обработки данных, который используется в машинном обучении. Она позволяет улучшить стабильность и скорость обучения, делая значения данных более сопоставимыми [21]. Для эффективного анализа и прогнозирования неоднородных изменений во времени на временных рядах может потребоваться приведение значений к определенному диапазону, например, от 0 до 1, путем масштабирования на основе минимальных и максимальных значений. Диапазон значений может быть различным для разных временных рядов, поэтому важно учитывать это при анализе.

Одним из методов нормализации является метод Min-Max нормализация, который позволяет привести данные к диапазону значений от 0 до 1 по следующей формуле:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}},$$

где x – значение данных, x_{min} – минимальное значение во всем наборе данных, x_{max} – максимальное значение во всем наборе данных, x_{norm} – нормализованное значение.

Нормализация проведена с помощью объекта MinMaxScaler библиотеки Scikit-learn. Результат нормализации приведен на рис. 6.

```
scaler=MinMaxScaler(feature_range=(0,1))
scaler.fit(np.array(closedf).reshape(-1,1))
closedf_train = scaler.transform
(np.array(closedf['EMA']).reshape(-1,1))
closedf_train[0:10]
array([[0.06392557],
       [0.06977791],
       [0.06509604],
       [0.06665066],
       [0.06677191],
       [0.0685916 ],
       [0.07091369],
       [0.07363772],
       [0.07668328],
```

Рисунок 6 – Нормализация начальных элементов выборки

Применение метода Min-Max нормализации к временным рядам позволяет привести данные к единому диапазону значений, что упрощает их анализ.

E. Составление наборов для обучения

Задача прогнозирования закупочной цены обладает теми особенностями, которые делают целесообразным использование именно интеллектуальных нейросетевых методов моделирования и, в частности, небольшой размер набора данных; возможность адаптации модели при поступлении новых данных; наличие

большого количества позиций номенклатуры.

В машинном обучении случайное разделение потока данных на тестовую и обучающую выборки считается нормой, так как существует независимость между одним наблюдением и другим. Однако, это утверждение не относится к временным рядам данных.

Поэтому в данной задаче мы разделим сформированный датасет последовательно на 3 выборки:

– 90% данных будем использовать для обучения (train), итого 3643 записи;

– 10% для проверки (valid), итого 405 записей;

– данные за 10 последних дней будем использовать для тестирования (test), итого 10 записей.

Выводы

В статье обсуждены проблемы разработки аналитических подсистем закупок лекарств аптечной сети. Отмечена важность процесса предобработки данных для их последующего использования в качестве датасета задач машинного обучения.

В результате были реализованы такие задачи предварительной обработки данных, как, очистка данных, заполнение пропусков элементов ряда значениями, сглаживание временного ряда с целью устранения выбросов, нормализация переменных и получение набора данных для дальнейшего обучения и тестирования в машинных алгоритмах.

Литература

1. Мариничев, И. И. Применение имитационного моделирования для принятия решения по управлению закупками при децентрализованной схеме поставок товара / И. И. Мариничев, Д. И. Трачук, В. А. Светличная // Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2021). - Донецк: ДонНТУ, 2021. – С. 100-104.

2. Шумаева, Е.А. Применение технологий блокчейн и интернет вещей в системе управления цепями поставок / Е. А. Шумаева, Т. К. Левина // Ресурсосбережение. Эффективность. Развитие : Материалы VI Международной научно-практической конференции, Донецк, 29 октября 2021 года. – Донецк: Донецкий национальный технический университет, 2021. – С. 635-639.

3. Миронкина, С.В. Особенности управления запасами в аптечной сети / С. В. Миронкина, Е. А. Шумаева // Актуальные проблемы социально-экономического развития современного общества : сборник статей II международной заочной научно-практической конференции, Киров, 27 мая 2021 года. – Киров:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Кировский государственный медицинский университет» Министерства здравоохранения Российской Федерации, 2021. – С. 289-293.

4. Светличная В.А. Использование методов теории принятия решений для выбора оптимальной стратегии при закупке лекарственных средств / В.А. Светличная, Е.А. Шумаева, О.В. Ченгарь, А.В. Андриевская // Экономика строительства и городского хозяйства. 2020. Т. 16. № 1. С. 41-48.

5. Андриевская, А.В. Экстраполяционные методы прогнозирования закупочных цен лекарств в условиях аптечной сети/ А.В. Андриевская, В.А. Светличная // В сборнике: Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2022). Материалы XIII Международной научно-технической конференции в рамках VIII Международного Научного форума Донецкой Народной Республики. Донецк, 2022. С. 195-201.

6. Золотова Ирина Юрьевна, Дворкин Владимир Валентинович Краткосрочное прогнозирование цен на российском оптовом рынке электроэнергии на основе нейронных сетей // Проблемы прогнозирования. 2017. №6.

7. Вовченко, В. О. Формирование датасета для решения задач машинного обучения / В. О. Вовченко, В. А. Светличная, Н. К. Андриевская // Информатика и кибернетика. – Донецк: ДонНТУ, 2023. – № 2(32). – С. 5-12.

8. Sergii K., Yurii S., Tatyana V., Natalia A. (2016) Feature Selection for Time-Series Prediction in Case of Undetermined Estimation. In: Samsonovich A., Klimov V., Rybina G. (eds) Biologically Inspired Cognitive Architectures (BICA) for Young Scientists. Advances in Intelligent Systems and Computing, vol 449. Springer, Cham

9. Stock prices dynamics forecasting with recurrent neural networks. Vasyaeva T., Martynenko T., Khmilovyi S., Andrievskaya N. // Открытые семантические технологии проектирования интеллектуальных систем, 2020. – № 4. – С. 277-282

10. Что такое предварительная обработка данных? - определение из техопедии [Электронный ресурс] – URL <https://ru.theastrologypage.com/data-preprocessing/> (дата обращения: 22.05.2023).

11. Прогнозирование временных рядов криптовалют для чайников [Электронный ресурс] – URL <https://vc.ru/u/1389654-machine-learning/580239-prognozirovanie-vremennyh-ryadov-kriptovalyut-dlya-chaynikov> // (дата обращения: 22.05.2023).

12. Интерполяция [Электронный ресурс] – URL: <https://wiki.loginom.ru/articles/interpolation.ht>

ml // (дата обращения: 21.05.2023).

13. Что такое Google Colab и кому он нужен [Электронный ресурс] – URL: <https://clck.ru/34Ypj3> // (дата обращения: 21.05.2023).

14. NumPy [Электронный ресурс] – URL <https://clck.ru/34YpjK> // (дата обращения: 22.05.2023).

15. Pandas [Электронный ресурс] – URL <https://blog.skillfactory.ru/glossary/pandas/> // (дата обращения: 22.05.2023).

16. Matplotlib [Электронный ресурс] – URL <https://clck.ru/34Ypju> // (дата обращения: 22.05.2023).

17. Scikit-learn [Электронный ресурс] – URL <https://clck.ru/34YpkF> // (дата обращения: 22.05.2023).

18. Keras [Электронный ресурс] – URL

<https://blog.skillfactory.ru/glossary/keras/> // (дата обращения: 22.05.2023).

19. Использование интерполяции для заполнения пропущенных записей в Python [Электронный ресурс] – URL <https://pythobyte.com/interpolation-to-fill-missing-entries-2ebaff45/> // (дата обращения: 22.05.2023).

20. Методы и модели сглаживания временных рядов [Электронный ресурс] – URL https://studme.org/209530/ekonomika/metody_model_i_sglazhivaniya_vremennyh_ryadov // (дата обращения: 22.05.2023).

21. Эксперименты с нейросетями (Часть 5): Нормализация входных параметров для передачи в нейросеть [Электронный ресурс] – URL <https://www.mql5.com/ru/articles/12459/> // (дата обращения: 22.05.2023).

Андриевская Н.К., Мартыненко Т.В. Разработка алгоритмов предобработки информации для прогнозных моделей ИСППР управления закупками. Статья посвящена разработке алгоритмов предварительной обработки данных и решению таких задач, как устранение неполноты данных и преобразование необработанной информации в удобный для анализа вид. При получении реальных данных временных рядов мы сталкиваемся с тем, что они лишь изредка бывают в необходимом формате. Существенная часть необработанных наборов данных слабо структурирована или содержит большое количество отсутствующих значений. Это делает невозможным последующую обработку данных некоторыми методами. В связи с этим реализованы основные этапы предварительной обработки данных, такие как, очистка данных, заполнение недостающих значений, сглаживание временного ряда, нормализация переменных. Таким образом, получен набор данных для дальнейших исследований, реализации и тестирования прогнозных моделей ИСППР управления закупками.

Ключевые слова: прогнозирование, лекарство, цена, закупка, временные ряды, предобработка.

Andrievskaya N.K., Martynenko T.V. Development of preprocessing information algorithms for predictive models for IDSS in procurement management. The article is devoted to the development of algorithms for preprocessing data and solving problems such as eliminating data incompleteness and converting raw information into a form convenient for analysis. When obtaining real time series data, we are faced with the fact that they only occasionally appear in the required format. A significant part of the raw data sets is poorly structured or contains a large number of missing values. This makes it impossible to process the data later by some methods. In this regard, the main stages of data preprocessing are implemented, such as data cleaning, filling in missing values, smoothing the time series, normalization of variables. Thus, a set of data was obtained for further research, implementation and testing for predictive models for IDSS in procurement management.

Keywords: forecasting, medicine, price, procurement, time series, preprocessing.

Статья поступила в редакцию 29.09.2023
Рекомендована к публикации доцентом Карабчевским В. В.