

Аннотация

Обнаружение неисправности сервисов продаж методом детекции аномалий во временных рядах

Юрченко Петр Олегович

В данной работе решается задача поиска неисправностей в сервисах продаж билетов и доп. услуг авиакомпании путем детекции аномалий во временных рядах, как обычных, так и сильно разреженных – с большим количеством нулей. Неотрицательность продаж накладывает на данные ограничение снизу, которое должны учитывать модели машинного обучения, направленные на поиск аномально низких показателей. Предложенный в работе подход направлен на то, чтобы показать, что качество работы FBProphet при использовании неотрицательных гамма и отрицательного биномиального распределения вместо нормального в основе модели с такого рода временными рядами выше, чем при использовании базовой реализации FBProphet. Также исследуется зависимость качества представленных моделей от ширины доверительного интервала и сравниваются несколько способов сглаживания данных скользящим средним.

Оглавление

1 Введение	2
2 Бизнес постановка задачи	3
2.1 Разнообразие сервисов продаж в авиакомпании	3
2.2 Выявление неисправностей в работе сервисов	3
2.3 Данные	4
3 Математическая постановка задачи	5
3.1 Аномалии во временных рядах	5
3.2 Создание искусственных аномалий	7
3.3 Классификация временных рядов	8
4 Критерии хорошего решения	11
5 Обзор существующих методов решения задачи	12
5.1 Методы поиска точечных аномалий в одномерных временных рядах . .	12
5.1.1 Методы, основанные на построении модели	12
5.1.2 Методы, основанные на плотности	15
5.1.3 Методы, основанные на гистограммном представлении	15
5.2 Методы поиска аномалий в разреженных временных рядах	16
6 Решение задачи	18
6.1 Используемые модели	18
6.1.1 Prophet	19
6.2 Выявление недостатков baseline решения	23
6.2.1 Внедрение других распределений в реализацию Prophet	23
6.3 Подбор ширины доверительного интервала	28
6.4 Добавление слаживания временного ряда	32
7 Анализ полученных результатов	37
8 Заключение	38

Глава 1

Введение

Временные ряды являются одними из самых распространенных видов данных для обработки. Чаще всего крупные компании для улучшения показателей исследуют неотрицательные временные ряды, такие как, данные о продажах, трафике, количестве пользователей или товаров и т. д., то есть ряды, значения которых по своей природе не могут быть ниже нуля. Одной из задач исследования временных рядов является поиск аномалий. Поиск аномалий направлен на обнаружение неожиданных/нежелательных событий или редких элементов в данных. Способность хорошо находить аномалии может привести к быстрому устранению неполадок, помочь избежать потери доходов и сохранить репутацию и бренд компании.

В данной работе мы решаем задачу детекции аномалий во временных рядах ежедневных продаж билетов и доп. услуг крупной авиакомпании для более быстрого обнаружения неисправностей в работе ее сервисов. В данном случае аномалия – нулевые продажи, то есть отсутствие возможности покупки с сервиса. Основная проблема, с которой сталкивается компания – таких сервисов продаж огромное количество. Клиент, покупая какую-то доп. услугу на определенное направление уже использует независимый, отдельно поддерживаемый сервис. Соответственно, мы хотели бы обрабатывать временные ряды с каждого отдельного направления, даже самого немногочисленного по продажам. На этом шаге мы сталкиваемся с необходимостью искать аномалии в разреженных временных рядах – рядах с большим количеством нулевых значений. Эта задача является наиболее трудоемкой и имеет не так много решений в литературе. Именно эта задача и будет основной целью исследования.

Для построения решения данной задачи за основу была взята модель FBProphet и две ее модификации с использованием неотрицательных распределений вместо нормального: гамма и неотрицательного биномиального распределения. В этой работе мы проверяем гипотезу, что данные модификации за счет свойств представленных распределений смогут хорошо аппроксимировать оклонулевые временные ряды, а также строить доверительные интервалы для детекции аномально низких продаж.

Итоговой целью работы является сравнение результатов, полученных от трех разновидностей FBProphet'a для различных типов временных рядов по критерию количества продаж и для разных временных интервалов аномалий, а также ответ на вопрос: сколько должен не работать сервис, чтобы модель обнаружила неисправность?

Глава 2

Бизнес постановка задачи

2.1 Разнообразие сервисов продаж в авиакомпании

Каждый день авиакомпания продает огромное количество билетов и доп. услуг клиентам. Эти продажи могут осуществляться через множество различных сервисов, таких как, сайт авиакомпании, мобильные приложения на разных девайсах, партнерские сервисы и так далее.

Разработка, обновление и поддержка всех этих сервисов является раздельной. Это значит, что добавление очередного обновления, новой маркетинговой стратегии или какого-либо кардинального изменения в любой из этих сервисов может привести к его неисправности. В данной работе мы будем рассматривать неисправности сервисов, которые невозможно обнаружить из логов, но которые для клиентов делают невозможной покупку на этих сервисах.

Нужно также понимать, что определенное обновление может касаться только конкретных направлений, по которым летает авиакомпания. То есть ситуация с остановкой продаж только по одному направлению вполне реальная. Такую неисправность не получится увидеть из временного ряда суммарных продаж, потому что она может иметь очень маленький вклад в него. Поэтому каждое направление авиакомпании необходимо рассматривать, как отдельно работающий сервис, в котором важно уметь находить неисправности.

Подводя итог вышесказанному, мы видим, что в компания ежедневно обслуживает огромное количество сервисов продаж, за которыми достаточно сложно одновременно следить, и, когда один из этих сервисов перестает работать по каким-либо причинам, не всегда это можно сразу заметить.

2.2 Выявление неисправностей в работе сервисов

Теперь, понимая проблематику, перейдем к постановке задачи. Мы хотим находить неисправности в работе сервисов пользуясь только статистикой продаж билетов или доп. услуг. Другими словами, задача заключается в обнаружении аномально низких продаж.

Помимо выявления неисправностей, решение могло бы служить хорошим инструментом для получения дополнительных сведений о поведении клиентов, оценке качества новой маркетинговой стратегии или чистки данных от выбросов для дальнейшего их использования в моделях.

2.3 Данные

Для исследования использовались временные ряды продаж билетов и доп. услуг авиакомпанией за каждый день в промежутке с февраля по ноябрь 2021 года включительно. Как было сказано выше, мы бы хотели уметь находить аномалии на отдельных направлениях и для отдельно взятых доп. услуг, поэтому все временные ряды будем смотреть как в общем, так и в разрезе отдельных направлений.

В авиакомпании периодически проходят распродажи, когда клиентский спрос ведет себя иначе. Поэтому в предоставленных данных явно заметны скачки в продажах билетов или наоборот спады в продажах некоторых доп. услуг в эти периоды. При решении задачи мы также хотели бы учитывать это.

Глава 3

Математическая постановка задачи

3.1 Аномалии во временных рядах

Рассматриваемые данные продаж – временные ряды. В нашем случае задача поиска неисправностей сервиса равносильна распространенной в литературе задаче поиска аномалий или выбросов во временной ряду.

Ознакомимся с терминологией задачи. Методы поиска аномалий во временных рядах сильно зависят от входных данных, типа искомой аномалии и от основы метода [1]. В этой секции мы постараемся покрыть большую часть случаев и понять, к какому из них относится наша задача. На рис. 3.1 указана классификация этих параметров.

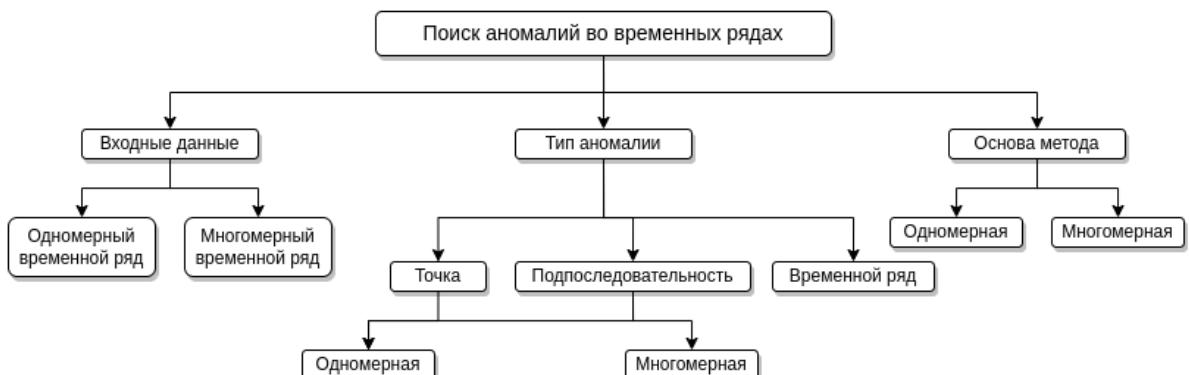


Рис. 3.1: Классификация аномалий

Далее опишем каждый из параметров отдельно. Все последующие определения взяты из [1].

Входные данные.

Из схемы на рис. 1 мы видим, что данные могут представлять собой одномерные и многомерные временные ряды.

Определение 3.1.1. Одномерный временной ряд $X = \{x_t\}_{t \in T}$, $x_t \in \mathbb{R}$ - это упорядоченный набор наблюдений, где каждое наблюдение получено в определенный момент времени $t \in T \subseteq \mathbb{Z}^+$.

x_t называется точкой или наблюдением, полученным в момент времени t , $S = x_p, x_{p+1}, \dots, x_{p+n-1}$ называется подпоследовательностью длины $n \leq |T|$, начинающейся с позиции p одномерного временного ряда X , где $p, t \in T$ and $p \leq |T| - n + 1$.

3.1. АНОМАЛИИ ВОВРЕМЕННЫХ РЯДОВ МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ

Предполагается, что каждое наблюдение x_t является реализацией случайной величины X_t .

Определение 3.1.2. *Многомерный временной ряд $\mathbf{X} = \{\mathbf{x}_t\}_{t \in T}, x_t \in \mathbb{R}^k$ - это упорядоченный набор векторов размерности k , каждый из которых получено в определенный момент времени $t \in T \subseteq \mathbb{Z}^+$ и состоит из k действительных наблюдений $\mathbf{x}_t = (x_{1t}, \dots, x_{kt})$.*

\mathbf{x}_t называется точкой или наблюдением, $S = x_p, x_{p+1}, \dots, x_{p+n-1}$ называется подпоследовательностью длины $n \leq |T|$ многомерного временного ряда \mathbf{X} , где $p, t \in T$ and $p \leq |T| - n + 1$. Для каждой размерности $j \in \{1, \dots, k\}$, $X_j = \{x_{jt}\}_{t \in T}$ - это одномерный временной ряд и каждое его наблюдение x_{jt} в векторе \mathbf{x}_t является реализацией случайной величины X_{jt} в $\mathbf{X}_t = (X_{1t}, \dots, X_{kt})$. В данном случае, каждая случайная величина зависит не только от прошлых ее реализаций, но также и от других случайных величин.

Тип аномалии.

Второй параметр описывает тип выброса, то есть то, что наш метод должен находить

- *Точка.* Точечный выброс – это данные, которые являются необычными в определенный момент времени по сравнению либо с другими значениями временного ряда (глобальный выброс), либо с соседними точками (локальный выброс). Точечные выбросы могут быть одномерными или многомерными в зависимости от того, влияют ли они на одну или несколько размерностей соответственно.
(можно привести пример картинкой)
- *Подпоследовательность.* Этот термин относится к последовательным моментам времени, совместное поведение которых является необычным, хотя каждое наблюдение в отдельности не обязательно является аномалией. Аномальные подпоследовательности также могут быть глобальными или локальными и могут влиять на одну или несколько размерностей.
- *Временной ряд.* Целые временные ряды также могут быть выбросами, но их можно обнаружить только тогда, когда входные данные представляют собой многомерные временные ряды.

Обратите внимание, что этот параметр тесно связан с типом входных данных. Если метод допускает только одномерные временные ряды в качестве входных данных, то многомерные аномальные точки или подпоследовательности не могут быть идентифицированы. Кроме того, аномальные временные ряды можно найти только в многомерных рядах. Наконец, следует отметить, что выбросы зависят от контекста. Таким образом, если метод обнаружения использует весь временной ряд в качестве контекстной информации, то обнаруженные выбросы являются глобальными. В противном случае, если метод использует только сегмент ряда (временное окно), обнаруженные выбросы являются локальными, поскольку они являются выбросами в своей окрестности. Глобальные выбросы также являются локальными, но не все локальные выбросы являются глобальными. Другими словами, некоторые локальные выбросы могут казаться нормальными, если наблюдать весь временной ряд, но могут быть аномальными, если мы сосредоточимся только на их окрестностях.

3.2. СОЗДАНИЕ ИСКУССТВЕННЫХ АНОМАЛИЙ

Основа метода.

Третий параметр описывает характер используемого метода обнаружения.

Одномерный метод обнаружения рассматривает только одну размерность, зависящую от времени, тогда как многомерный метод обнаружения может одновременно работать с более чем одной размерностью, зависящей от времени. Обратите внимание, что метод обнаружения может быть одномерным, даже если входные данные представляют собой многомерный временной ряд, потому что индивидуальный анализ может быть выполнен для каждой размерности, без учета зависимостей, которые могут существовать между остальными размерностями. Напротив, многомерный метод нельзя использовать, если входные данные представляют собой одномерный временной ряд. Следовательно, этот параметр будет рассматриваться только для данных многомерных временных рядов.

В нашей задаче входные данные – одномерный временной ряд, описывающий количество продаж каждый день с одного конкретного сервиса. Тип аномалии: нам подходит точка или подпоследовательность в зависимости от того, сколько дней сервис не работал. В этой работе будем считать, что ищем выброс из данных в виде точки, и если метод находит хотя бы одну точку в аномальной подпоследовательности, то находит и всю подпоследовательность. Этого будет достаточно для детекции неисправности самого сервиса. Основа метода в нашем случае может быть только одномерной, потому что рассматривается одномерный временной ряд. Основные методы детекции конкретно таких выбросов описаны в пункте 5.1.

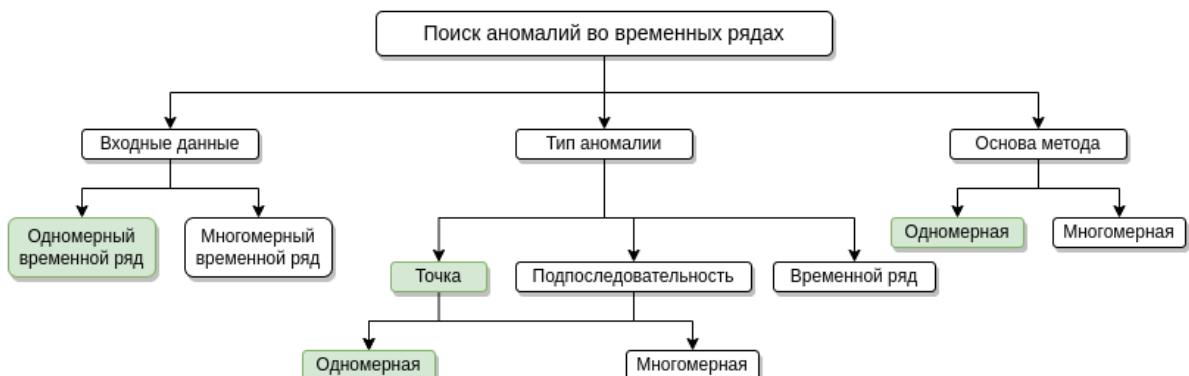


Рис. 3.2: Классификация аномалий

3.2 Создание искусственных аномалий

Сейчас, определившись с тем, что мы решаем задачу поиска аномалий во временных рядах, нужно определить, как мы будем оценивать качество нашего решения. Так как мы имеем дело с историческими данными, понять работал сервис в определенный день или не работал невозможно, поэтому будем искусственно создавать аномалии различной продолжительности в наших временных рядах. Другими словами, будем “выключать” наш сервис в прошлом на определенное время и смотреть, сможет ли модель обнаружить данную аномалию.

3.3. КЛАССИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ САЯ ПОСТАНОВКА ЗАДАЧИ

Были выбраны следующие временные интервалы для создания аномалий:

1. 4 часа
2. 8 часов
3. 24 часа
4. 72 часа

После того, как мы выбрали временной ряд для анализа, случайнм образом выбирается день и час, в котором будет “выключен” сервис продаж. После этого момента значение покупок будет равно нулю на протяжении указанных выше временных интервалов. Важно, что в такой реализации аномалия может быть переходящей и не сильно выраженной. То есть если “выключение” сервиса длительностью 8 часов начнется в 20:00 одного дня, то будет закончено в 4:00 следующего дня и значения покупок изменятся незначительно - всего на 1/6 от общего числа и это может быть незаметно для модели. Такая реализация создания аномалий была выбрана, как наиболее близкая к реальности. Но для модели такие аномалии будет сложнее находить. Будем считать, что модель обнаружила всю аномалию, если обнаружила хотя бы один день, входящий в интервал аномалии.

Для оценки качества модели на определенном временном ряде и с определенной длиной аномалии будем проводить 80 экспериментов с созданием искусственной аномалии в случайном месте ряда, как было описано выше. Далее будем считать количество правильных и ложных срабатываний модели и рассчитывать метрики precision, recall и F1. Критерии успеха модели описаны в пункте 4.

3.3 Классификация временных рядов

В данной постановке задачи мы выделили 3 типа временных рядов с точки зрения сложности нахождения аномалий:

1. *Популярное направление* – временной ряд с большими значениями покупок каждый день:



Рис. 3.3: Пример популярного направления

3.3. КЛАССИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ САЯ ПОСТАНОВКА ЗАДАЧИ

2. *Непопулярное направление* – временной ряд с небольшими, но ненулевыми значениями покупок каждый день:



Рис. 3.4: Пример непопулярного направления

3. *Направление с почти нулевым спросом* – временной ряд с большим количеством нулевых покупок каждый день:



Рис. 3.5: Пример направления с почти нулевым спросом

В исследовании мы не можем рассмотреть все возможные временные временные ряды, поэтому выберем несколько каждого типа. Будем решать задачу для следующих шести рядов, с учетом, что решение можно обобщить на все остальные ряды:

1. Популярное направление:
 - 1.1. общий факт покупок доп. услуг
 - 1.2. услуга выбора места в самолете
2. Непопулярное направление:
 - 2.1. подарочный сертификат
3. Направление с почти нулевым спросом:
 - 3.1. направление 'AAQ-DME', услуга автоматической регистрации на рейс
 - 3.2. направление 'DME-GDZ', услуга автоматической регистрации на рейс

3.3. КЛАССИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ САЯ ПОСТАНОВКА ЗАДАЧИ

3.3. направление 'DME-ORY', услуга выбора места

Для каждого из этих трех типов нужен свой подход к нахождению в нем аномалий. Если с первым типом никаких проблем нет, то со вторым и третьим сразу видно, что не все так просто. Обычные модели, аппроксимирующие временной ряд, не учитывают, что отрицательные значения недопустимы из-за физики процесса.

С этой проблемой мы сталкиваемся и в нашей задаче. Продажи не могут быть отрицательными, и нулевые показатели зачастую уже являются аномальными. При этом, используя модель можно спокойно получать отрицательные значения, не говоря уже о получаемом доверительном интервале модели.

Таким образом, видим, что основная трудность задачи заключается именно в рассмотрении 2-го и 3-го типов временных рядов. Решение данной проблемы и будет основной целью этой работы.

Глава 4

Критерии хорошего решения

В пункте (3.2) описано, как проводились эксперименты и рассчитывались метрики качества модели в определенном временном ряде.

Напомню, что мы проводили по 80 экспериментов с искусственным “выключением” сервиса на 4, 8, 24 и 72 часа и, подсчитывали метрики, считая, что если модель нашла один день из аномалии, то она нашла всю аномалию:

$$\text{Precision} = \frac{tp}{tp + fp} \quad (4.1)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (4.2)$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.3)$$

Критерии:

1. Целью задачи является нахождение фактов неполадок в работе сервисов продаж, поэтому нас интересуют только аномально низкие значения, аномально высокие мы не будем брать в рассмотрение
2. Мы хотим, чтобы модель не пропускала аномалии - детектировала все или большинство случаев отключения сервиса. То есть хотим иметь высокий recall модели. Добиться $\text{recall} = 1$ достаточно сложно для большинства случаев, поэтому потребуем, чтобы модель достигала $\text{recall} > 0.85$
3. Очевидно, что если взять самый узкий доверительный интервал модели для предсказаний, то мы получим максимальный recall, но тогда все остальные точки временного ряда, которые лежат ниже предсказания будут также обозначены аномальными, что не несет никакой смысловой нагрузки. Также мы не хотим вручную проверять слишком много положительных решений модели, поэтому для нас также будет важно смотреть на метрику precision. Потребуем, чтобы на 3 положительных предсказания модели мы имели точно одну аномалию, то есть $\text{precision} > 0.33$
4. Таким образом требуем, чтобы F1 был больше 0.56

Глава 5

Обзор существующих методов решения задачи

5.1 Методы поиска точечных аномалий в одномерных временных рядах

Методы, которые будут обсуждаться в этом разделе, предназначены для обнаружения точечных выбросов в одномерном временном ряду. В пункте 3.1 мы показали, что именно такая задача равносильна нашей бизнес постановке.

Основные подходы представлены в виде диаграммы, показанной на рис. 5.1. Напомним, что для обнаружения одномерных выбросов можно использовать только методы одномерного обнаружения [1].

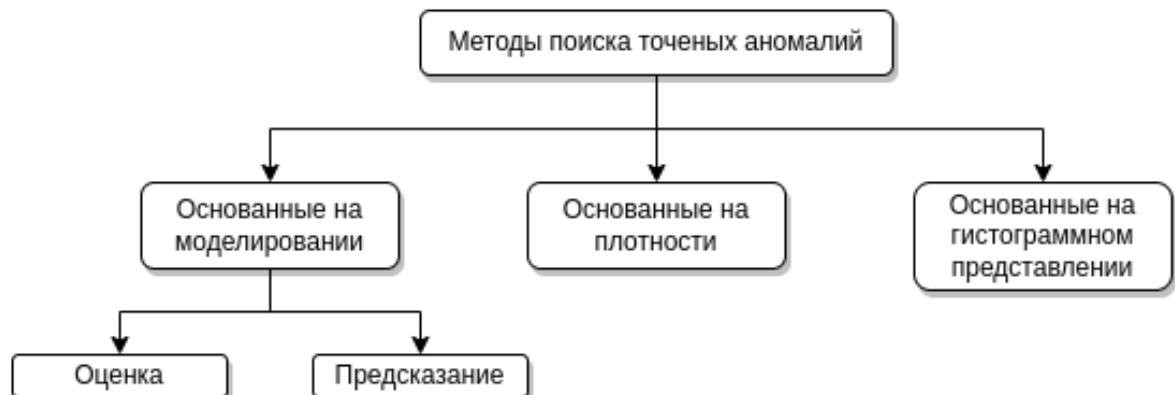


Рис. 5.1: Методы поиска аномалий

5.1.1 Методы, основанные на построении модели

Определение 5.1.1. Точечная аномалия - это точка, которая значительно отклоняется от ожидаемого значения.

Учитывая одномерный временной ряд в нашей задаче, точка во время t может быть объявлена выбросом, если расстояние до ее ожидаемого значения выше заранее определенного порога τ :

$$|x_t - \hat{x}_t| > \tau \quad (5.1)$$

5.1. МЕТОДЫ ПОИСКА ТОЧЕЧНЫХ АНОМАЛИЙ В ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДАХ ЗАДАЧИ

где x_t - наблюдаемая точка данных, а \hat{x}_t - ее ожидаемое значение.

Методы обнаружения выбросов, основанные на стратегии, описанной в неравенстве (5.1), называются *методами на основе моделирования* и являются наиболее распространенными подходами в литературе. Несмотря на то, что каждый метод вычисляет ожидаемое значение \hat{x}_t и порог τ по-разному, все они основаны на обучении модели. Как показано в таблице 5.1, если \hat{x}_t получается с использованием предыдущих, и последующих наблюдений за x_t (прошлые, текущие и будущие данные), то этот метод называется основанным на оценочном моделировании. Напротив, если \hat{x}_t получается, полагаясь только на предыдущие наблюдения x_t (прошлые данные), то этот метод называется основанным на прогнозном моделировании. На практике основное различие между использованием методов оценки и прогнозирования заключается в том, что методы из последней категории могут использоваться в потоковых временных рядах, поскольку они могут определять, является ли новое значение выбросом, как только оно поступает. В случае оценочного моделирования это можно сделать только в том случае, если помимо точек, предшествующих x_t , для вычисления оцениваемого значения используется еще и текущая точка $x_t (k_2 = 0)$.

Таблица 5.1: Используемые данные для методов, основанных на моделировании для $k_1, k_2 \geq 0$, так что $k_1 + k_2 > 0$.

	Используемые данные	Ожидаемое значение	Выбросы
Оценочное моделирование	$\{x_{t-k_1}, \dots, x_t, \dots, x_{t+k_2}\}$	\hat{x}_t	$ x_t - \hat{x}_t > \tau$
Предсказательное моделирование	$\{x_{t-k_1}, \dots, x_{t-1}\}$	\hat{x}_t	$ x_t - \hat{x}_t > \tau$

Самые простые модели оценки основаны на постоянных или кусочно-постоянных моделях, где базовые статистики, такие как медиана [2] или медианное абсолютное отклонение (MAD) [3] используются для получения ожидаемого значения \hat{x}_t . Эти статистики рассчитываются с использованием полного ряда или путем группировки данных в сегменты равной длины. Более сложный подход заключается в использовании сегментов неравной длины, полученных при помощи сегментации. В [4] используют среднее значение каждого сегмента для определения ожидаемого значения ряда в этом сегменте. В статье также используют адаптивный порог $\tau = \alpha\sigma_i$, где α - фиксированное значение, а σ_i - стандартное отклонение сегмента i .

Другие методы, основанные на оценке, предназначены для нахождения точек, которые маловероятны, если предполагается, что данные были сгенерированы определенной моделью или распределением. Например, некоторые авторы моделируют структуру данных, используя методы сглаживания, такие как B-сплайны и ядра [5] или варианты экспоненциально взвешенного скользящего среднего (EWMA) [6]. В [7] и [8] авторы используют ограничения по уклону между точками в данных, в [9] - предполагают, что данные распределены нормально, если не учитывать выбросы, а в [10] используют гауссово смешанные модели (GMM).

Как только модель или распределение подобраны и обучены, в [5], [6] и [10] используют неравенство (5.1) напрямую, чтобы решить, является ли точка аномалией или нет. Точно так же в статье [7] устанавливают максимально и минимально возможный наклон между последовательными значениями, тогда как в [8] моделируют изменение уклонов до и после момента времени t , предполагая, что уклоны не должны существенно меняться в момент времени. Наконец, в [9] для принятия решения проверяется тест Extreme Studentized Deviate (ESD). Его нулевая гипотеза состоит в том, что выбросов нет, а альтернатива - в том, что их количество может достигать k .

5.1. МЕТОДЫ ПОИСКА ТОЧЕЧНЫХ АНОМАЛИЙ В ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДАХ ЗАДАЧИ

Независимо от временной корреляции, алгоритм итеративно вычисляет k тестовых статистик, чтобы обнаружить k выбросов. На каждой итерации удаляется самое удаленное от среднего значения наблюдение и затем рассчитывается новая статистика.

Другие одномерные методы детекции выбросов анализируют уже разности фактического значения и предсказания, полученного из разных моделей, для выявления выбросов. Например, в [11] используют декомпозицию STL (Seasonal and Trend decomposition using Loess), а в [12], [13], [14] используют модели ARIMA, линейную регрессию и нейронные сети соответственно. Большинство из этих моделей можно использовать не только в оценочном, но также в предсказательном моделировании. После того, как выбранная модель обучена, проверка гипотез применяется к разностям действительного и предсказанного значений для обнаружения выбросов. В работах [12], [13], [14], предполагая, что распределение этих разностей известно, минимальное и максимальное значения проверяются одновременно на каждой итерации алгоритма. Проверяемая гипотеза заключается в том, является ли экстремум выбросом (альтернативная гипотеза) или нет (нулевая гипотеза). Обнаруженные выбросы корректируются к среднему значению, и процесс повторяется до тех пор, пока не будут обнаружены выбросы. Напротив, в [11] применяется тест ESD, описанный в предыдущем абзаце, но для надежности используются медиана и MAD, а не среднее значение и стандартное отклонение.

В отличие от оценочных моделей, методы, основанные на моделях прогнозирования, получают \hat{x}_t , используя только прошлые данные, то есть без использования текущей точки x_t или каких-либо апостериорных наблюдений. Точки, сильно отличающиеся от своих предсказанных значений, идентифицируются как выбросы. Напомним, что все методы этой категории могут работать с потоковыми временными рядами.

В рамках методов, основанных на прогнозировании, некоторые используют фиксированную модель и, таким образом, не могут адаптироваться к изменениям, происходящим в данных с течением времени. Например, алгоритм DeepAnT, представленный [15] – это основанный на глубоком обучении подход к обнаружению выбросов, который применяет сверточные нейронные сети (CNN) для прогнозирования значений в будущем. В других методах используется авторегрессионные модели, как в [16], модель ARIMA в [17] или модель Prophet в [18], основанная на декомпозиции, которые возвращают доверительные интервалы для прогнозов, а не только точечные оценки. Следовательно, эти методы неявно определяют значение τ , что делает эти методы более удобными в применении.

Также бывают методы, которые адаптируются к эволюции временных рядов путем переобучения модели с появлением новых точек. В качестве наиболее простого подхода в [19] описывается метод, который предсказывает значение \hat{x}_t с помощью медианы его прошлых данных. С поступлением новых точек во временной ряд можно достаточно быстро переобучать модель для предсказания будущих значений. В [20] используется более продвинутый метод: авторы обучают модель ARIMA на скользящем окне, и параметры обновляются каждый раз, когда окно перемещается на шаг вперед.

Теория экстремальных значений — это еще один подход, основанный на прогнозировании, который использовался для обнаружения точечных выбросов в потоковых одномерных рядах с использованием прошлых данных и повторного обучения. При фиксированном риске q эта теория позволяет нам получить пороговое значение $z_{q,t}$, которое адаптируется к эволюции данных, так что $P(X_t > z_{q,t}) < q$, предполагая, что экстремальные значения следуют обобщенному распределению Парето (GPD).

5.1. МЕТОДЫ ПОИСКА ТОЧЕЧНЫХ АНОМАЛИЙ В ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДАХ ЗАДАЧИ

Входящие данные используются как для обнаружения аномалий ($X_t > z_{q,t}$), так и для уточнения $z_{q,t}$. Этот метод допустим для нахождения верхних выбросов, идея аналогична для нахождения нижних выбросов или обоих. Эти подходы примечательны тем, что для нахождения аномалий нам не нужно делать предположение о распределении самих данных, а только о распределении выбросов. Методы, основанные на этой теории, описаны в [21] и [22]: SPOT для временных рядов, являющихся стационарными и DSPOT для данных, которые могут быть нестационарными.

В методах, основанных на прогнозировании, необходимо поддерживать актуальность данных, поэтому все описанные ранее подходы используют переобучение модели на новых точках либо с каким-то постоянным периодом, либо каждый раз, когда появляется новая точка. Однако, ни один из них не применяет инкрементальные подходы к обучению модели, когда модель не переобучается каждый раз с нуля, а итеративно обновляется, используя только новую полученную информацию. Это позволяет избежать затрат, связанных с обучением моделей более одного раза, и позволяет более плавно адаптировать модель к изменениям, которые могут произойти в данных, что представляет особый интерес в контексте потоковой передачи. В этом смысле и в отличие от предыдущих подходов в [23] и [24] предлагаются моделировать одномерный временной ряд пошагово. Этот метод использует распределение Стьюдента для вычисления доверительного интервала и обновляет ковариационную матрицу новой полученной точкой данных. В [25] используют сеть иерархической временной памяти (HTM), которая также представляет собой метод, основанный на моделировании, который постепенно обновляется по мере поступления новых наблюдений.

5.1.2 Методы, основанные на плотности

Все вышеописанные методы основаны на неравенстве (5.1). Однако, не все существующие методы обнаружения точечных выбросов основаны на этой идее. Например, далее мы рассмотрим методы на основе плотности, которые относятся ко второй категории методов, изображенных на рис. 5.1. Методы в этой группе предполагают, что точки с менее τ соседями являются выбросами. Здесь соседи - точки, которые находятся на расстоянии R от рассматриваемой точки. Это можно обозначить следующим образом:

Определение 5.1.2.

$$x_t - \text{аномалия} \iff |\{x \in X | d(x, x_t) \leq R\}| < \tau \quad (5.2)$$

где d - расстояние (чаще всего евклидово), x_t представляет собой анализируемую точку данных в момент времени t , X - набор точек данных и $R \in \mathbb{R}^+$. Таким образом, точка является выбросом, если $\tau_p + \tau_s < \tau$, где τ_p и τ_s - количество предшествующих и последующих соседей x_t на расстоянии меньшем или равном R .

Обнаружение выбросов на основе плотности широко используется в невременных данных, но концепция соседства более сложна во временных рядах, поскольку данные упорядочены. Чтобы принять во внимание эту зависимость от времени, в [26] и [27] применяют этот метод со скользящим окном, что позволяет нам определить, является ли значение ряда выбросом не только глобально, но и в окрестности.

5.1.3 Методы, основанные на гистограммном представлении

Рассмотрим последнюю группу, анализируемую в этом разделе. Этот тип метода основан на нахождении точек, удаление которых из одномерного временного

5.2. МЕТОДЫ ПОИСКА АНОМАЛИЙ В РАЗРЕЖЕННЫХ ВРЕМЕННЫХ РЯДАХ

ЛАВА 5. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ

ряда приводит к гистограммному представлению с меньшей ошибкой, чем исходная, даже после того, как количество сегментов было уменьшено для учета раздельного нахождения этих точек. Гистограмма строится путем вычисления среднего значения в каждом сегменте. Тогда

Определение 5.1.3. Для заданной последовательности точек X и числа сегментов гистограммы B множество $D \subset X$ является аномальным, если

$$E_X(H_B^*) > E_{X-D}(H_{B-|D|}^*) \quad (5.3)$$

где H_B^* - оптимальная гистограмма (гистограмма с наименьшей ошибкой аппроксимации) для данных X с B сегментами, $E_X()$ - общая ошибка аппроксимации, а $H_{B-|D|}^*$ - оптимальная гистограмма на $X - D$ с $B - |D|$ сегментах. В [28] был предложен механизм динамического программирования для построения гистограммы, состоящей из $B - |D|$ сегментов и $|D|$ аномалий, сводя к минимуму его общую ошибку. Несколько лет спустя, в [29] был представлен не только оптимальный алгоритм для непотоковых временных рядов, но также приближенный алгоритм и эвристический подход для потокового случая.

5.2 Методы поиска аномалий в разреженных временных рядах

Как мы видим, существует большое множество способов для нахождения аномалий во временных рядах, но, как уже было сказано, основной задачей работы является нахождение аномалий в неотрицательных рядах с большим количеством нулей. В этом пункте мы рассмотрим как раз те методы, которые используются при аппроксимации таких временных рядов.

Одним из первых подходов для предсказания разреженных временных рядов предложил Кростон в [30]. Так как временной ряд является чередованием нулевых и ненулевых точек, то было предложено изучать два объекта: вероятность того, что спрос будет ненулевым, и величину этого ненулевого спроса.

Подход Кростона является итеративным и состоит из трех шагов:

- Оценка среднего спроса за период

Вводится оценочный уровень спроса a и актуальный уровень спроса d , после чего оценивается уровень спроса на следующем шаге:

$$a_{t+1} = \begin{cases} \alpha d_t + (1 - \alpha)a_t, & \text{if } d_t > 0 \\ a_t, & \text{if } d_t = 0 \end{cases}. \quad (5.4)$$

где $\alpha \in (0, 1)$ – параметр обучения, которое отвечает за важность самого последнего события по отношению к историческим.

- Оценка среднего времени между двумя моментами возникновения спроса

Аналогично прошлому пункту вводится оценочное время между двумя ненулевыми событиями - периодичность p и количество нулей, начиная с прошлого ненулевого спроса q , после чего обновляется p на следующем шаге. Следующее обновление параметров происходит, только если спрос не текущем шаге ненулевой.

$$p_{t+1} = \begin{cases} \alpha q + (1 - \alpha)p_t, & \text{if } d_t > 0 \\ p_t, & \text{if } d_t = 0 \end{cases} . \quad (5.5)$$

где α – тот же самый параметр.

- Прогнозирование спроса, как произведение вероятности того, что спрос будет ненулевым и ранее оцененного уровня спроса

$$f_{t+1} = \frac{a_t}{p_t} \quad (5.6)$$

Но у метода Кронсона есть один один недостаток – когда нет ненулевых наблюдений, прогноз никак не обновляется. В [31] строится модификация подхода Кронсона – модель TSB. В ней оценочный уровень спроса a остается таким же, как в исходной модели, новизной является изменение параметра периодичности p . Теперь эта величина будет означать вероятность появления ненулевого спроса и будет обновляться на каждом шаге, а не только, когда есть спрос:

$$p_{t+1} = \begin{cases} \beta + (1 - \beta)p_t, & \text{if } d_t > 0 \\ (1 - \beta)p_t, & \text{if } d_t = 0 \end{cases} . \quad (5.7)$$

где β – аналогичный α параметр.

И, прогноз теперь будет равен $f_{t+1} = p_{t+1} \cdot a_{t+1}$. В такой реализации удалось получить логичное уменьшение прогноза, когда нет никакого спроса.

Другой подход для предсказания разреженных временных рядов был представлен в [32]. Чтобы уйти от сильной разреженности, авторы добавили агрегирование данных скользящим окном, таким образом получив уже отличный от нуля временной ряд, который лучше поддается предсказанию. Затем были использованы статистические методы перехода от агрегированных предсказаний в обратную сторону и получена зависимость качества от ширины скользящего окна.

В [33], [34] и [35] представлено сравнение работы моделей Cronson, TSB, ARIMA, SARIMA, Prophet и разных вариаций LSTM'а для предсказания неотрицательных близких к нулю временных рядов и показано, что некоторые нейронные сети чуть лучше справляются с задачей аппроксимации, чем классические модели. Также в [34] показано, что для прогнозирования продаж по качеству сразу после LSTM идет модель Prophet. Однако, эти работы не нацелены конкретно на поиск аномальных значений, в них рассматривается именно аппроксимация временных рядов.

Глава 6

Решение задачи

6.1 Используемые модели

Из обзора литературы видно, что существует очень много различных методов поиска точечных аномалий во временных рядах. В данной работе мы не сможем покрыть их все, поэтому была выбрана одна категория методов: методы, основанные на оценочном моделировании рис. (6.1). Эта категория была выбрана по нескольким причинам:

- Наши данные имеют ярко выраженную сезонность и выбросы в дни распродажи, которые, нам необходимо учитывать при решении. В методах поиска точечных аномалий, основанных на плотности и гистограммном представлении, такой возможности нет, либо это было бы трудно имплементировать. Поэтому было принято решение остановиться на подходах, основанных на построении моделей, так как множество современных моделей умеют это делать “из коробки” и удовлетворяют нашим требованиям.
- В нашей задаче мы рассматриваем аномалии, которые уже произошли в прошлом, то есть методы прогнозного моделирования нам не подойдут

Соответственно, методы, основанные на оценочном моделировании - единственная подходящая для нашей задачи категория.

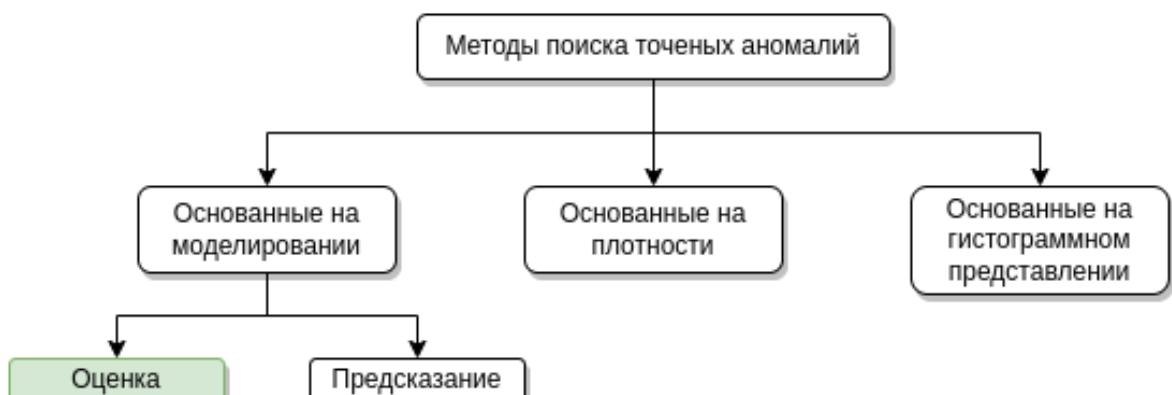


Рис. 6.1: Методы поиска аномалий

6.1.1 Prophet

Обзор модели

В 2017 году Facebook предложил модель Prophet [18], которая разработана на основе декомпозиции временных рядов. При обучении модели Prophet использует специальный инструмент для статистического моделирования pyStan [36], что позволяет получать предсказания быстрее конкурентных моделей. Модель Prophet чрезвычайно популярна в прогнозировании временных рядов из-за ее простой работы, легкой интерпретации параметров модели, небольших временных затрат и хороших результатов.

По своей сути Prophet представляет из себя сумму трех компонент, зависящих от времени – тренд $g(t)$, сезонность $s(t)$, праздники и события $h(t)$ и ошибки ε_t :

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (6.1)$$

Член ошибки ε_t представляет собой любые случайные флуктуации, которые не учитываются моделью. Prophet делает параметрическое предположение, что ε_t имеет нормальное распределение.

Функция тренда $g(t)$ имеет 3 различные опции представления:

1. Логистический рост:

Это представление полезно, когда временной ряд имеет верхний или нижний предел, в котором значения, которые вы моделируете, становятся насыщенными и не могут превзойти максимальное или минимальное значение. Когда используется это представление, тренд будет похож на типичное уравнение для логистической кривой

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (6.2)$$

где C - пропускная способность, k - скорость роста, а m - параметр сдвига.

Исследователи из Facebook пошли немного дальше и заменили пропускную способность C на зависимую от времени $C(t)$, а также постоянную k на изменяемую время от времени скорость роста следующим образом:

Было введено понятие точки изменения тренда – точка, в которой скорость роста меняет свое направление. Пусть явно заданы S точек изменения тренда в моменты времени $s_j, j = 1, \dots, S$. Определим вектор корректировок $\boldsymbol{\delta} \in \mathbb{R}^S$, где δ_j - изменение скорости в момент времени s_j . Тогда скорость роста в момент времени t равна базовой скорости k плюс все корректировки до этого момента: $k + \sum_{j:t>s_j} \delta_j$. Чтобы записать это в более простом виде введем вектор $\mathbf{a}(t) \in \{0, 1\}^S$ следующим образом

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j, \\ 0, & \text{otherwise} \end{cases} \quad (6.3)$$

Тогда скорость роста в момент времени t равна $k + \mathbf{a}(t)^T \boldsymbol{\delta}$. Когда k меняется в точке изменения тренда параметр сдвига также должен быть добавлен, чтобы соединить концы сегментов. Для корректного соединения сегментов в точке изменения j введем

$$\gamma_j = (s_j - m - \sum_{l < j} y_l) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l}\right) \quad (6.4)$$

Итоговая кусочно-логистическая модель роста имеет вид:

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \mathbf{a}(t)^T \boldsymbol{\delta})(t - (m + \mathbf{a}(t)^T \boldsymbol{\gamma})))} \quad (6.5)$$

2. Линейный рост:

Модель использует набор кусочно-линейных уравнений с разными наклонами между точками изменения тренда. Когда используется линейный рост, член роста будет похож на классический $y = at + b$, за исключением того, что наклон a и смещение b являются переменными и будут изменять значение в каждой точке изменения тренда.

Записывая в терминах точек изменения тренда, получим:

$$g(t) = (k + \mathbf{a}(t)^T \boldsymbol{\delta})t - (m + \mathbf{a}(t)^T \boldsymbol{\gamma}) \quad (6.6)$$

3. Плоский тренд:

Наконец, вы можете выбрать плоскую тренд, когда нет роста с течением времени (но все же может быть сезонность). В этом случае функция роста будет иметь постоянное значение.

Функция сезонности $s(t)$ – это ряд Фурье.

$$s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P})) \quad (6.7)$$

Обучение функции сезонности требует оценки $2N$ параметров $\boldsymbol{\beta} = [a_1, b_1, \dots, a_N, b_N]^T$. Это делается за счет построения матрицы сезонности для каждого t . В Prophet берется $\boldsymbol{\beta} \sim \text{Normal}(0, \sigma^2)$.

$$s(t) = X(t)\boldsymbol{\beta} \quad (6.8)$$

Функция праздников и событий $h(t)$ позволяет Prophet корректировать прогноз, когда праздник или важное событие могут изменить прогноз. Модель принимает список дат, и когда дата присутствует в прогнозе, добавляет или вычитает значение из прогноза. Также есть возможность задать диапазон дней вокруг дат или использовать встроенные списки праздников.

Обучение очень похоже на процедуру с сезонностью. Создается матрица

$$Z(t) = [\mathbf{1}(t \in D_1), \dots, \mathbf{1}(t \in D_L)] \quad (6.9)$$

где D_i - дни праздников.

$$h(t) = Z(t)\boldsymbol{\kappa} \quad (6.10)$$

Как и с сезонностью, в библиотеке используется априорное распределение $\boldsymbol{\kappa} \sim \text{Normal}(0, \nu^2)$.

Применение модели

Для построения baseline решения FBProphet был выбран из множества моделей для аппроксимации временных рядов по следующим причинам:

- Из обзора выше видно, что модель предоставляет гибкий функционал для учета сезонности и распродаж, о которых мы говорили раньше
- Модель не требует от данных стационарности и дополнительной предобработки
- Для валидации мы проводим по 80 экспериментов для каждого отдельного случая. Таких случаев: 4 (временных интервала аномалии) · 6 (выбранных временных рядов) = 24, то есть в сумме обучаем модель 1920 раз. Быстрое обучение Prophet'a играет тут большую роль
- Результаты, полученные из модели легко интерпретировать из-за возможности посмотреть каждое слагаемое из формулы (6.1) отдельно
- Prophet предоставляет доверительный интервал “из коробки”, который можно использовать, как критерий для обнаружения аномалий

При обучении модели была оставлена только недельная сезонность, так как у нас данные собраны на временном промежутке меньше года и в них нет почасового разбиения.

Рассмотрим полученные результаты для разных типов временных рядов и разной длительности аномалий.

На рисунках (6.2) - (6.4) представлены примеры работы baseline модели на разных временных интервалах аномалий. Красными точками обозначены аномальные значения, которые были обнаружены моделью, зелеными точками - аномалии, которые не были обнаружены моделью, а желтыми - ложные срабатывания модели.

1. Популярное направление:

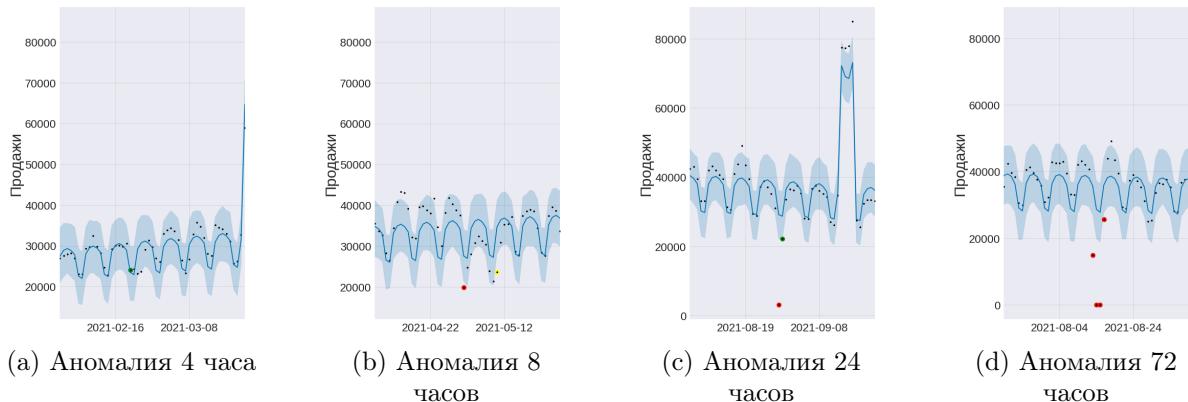


Рис. 6.2: Детекция аномалий baseline моделью на популярном направлении

Таблица 6.1: Метрики Baseline Prophet на популярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.09	0.23	0.3	0.73
Precision	0.05	0.13	0.18	0.57
Recall	0.31	0.85	1.0	1.0

2. Непопулярное направление:

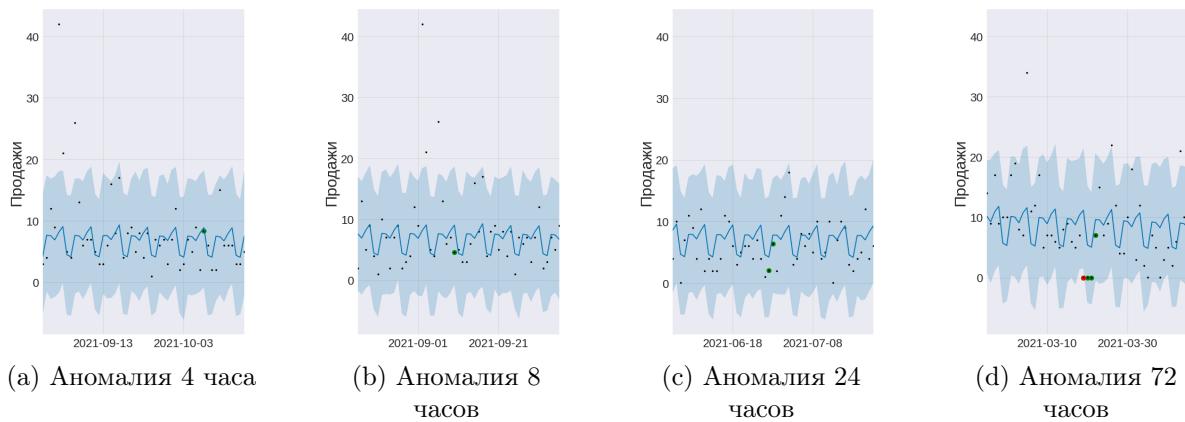


Рис. 6.3: Детекция аномалий baseline моделью на непопулярном направлении

Таблица 6.2: Метрики Baseline Prophet на непопулярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.0	0.0	0.01	0.2
Precision	0.0	0.0	1.0	1.0
Recall	0.0	0.0	0.05	0.11

3. Направление с большим количеством нулевых значений:

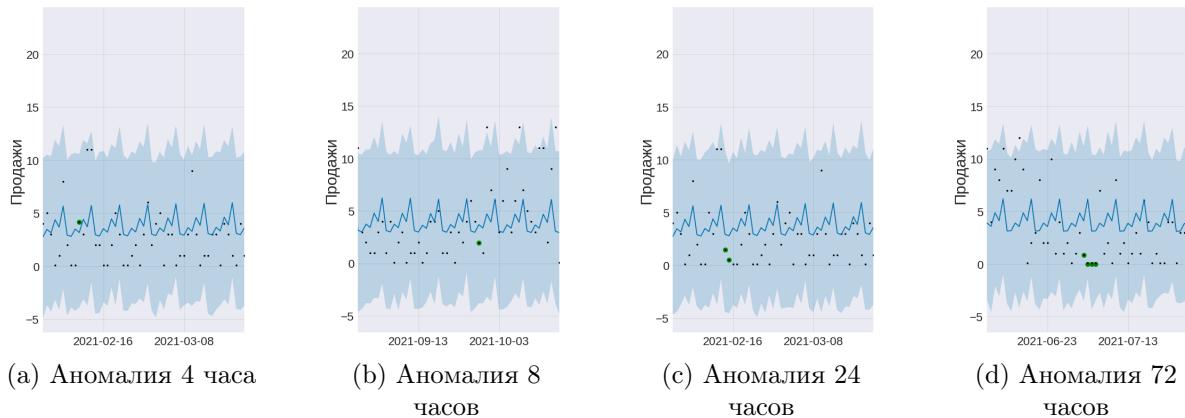


Рис. 6.4: Детекция аномалий baseline моделью на направлении с большим количеством нулей

Таблица 6.3: Метрики Baseline Prophet на направлении с большим количеством нулей

	4 часа	8 часов	24 часа	72 часа
F1	0.0	0.0	0.0	0.01
Precision	0.0	0.0	0.0	0.02
Recall	0.0	0.0	0.0	0.004

Видим, что данная модель хорошо справляется с решением задачи только в случае с популярным направлением для аномалии длительностью 3 суток. Для длительности аномалии 8 часов и 24 часов на популярном направлении получили ненулевые метрики, но все еще не дотягивающие до требуемых критериев (4); предполагаем, что подбор гиперпараметров для улучшения модели может помочь довести данные случаи до удовлетворяющих метрик. В других типах направлений Prophet неприменим для задачи, потому что нижняя граница доверительного интервала модели ниже нуля, а продажи всегда неотрицательны.

На самом деле, проблема попадания нулевых значений в доверительный интервал – это не только проблема Prophet'a, с данной проблемой столкнулась бы любая другая модель.

6.2 Выявление недостатков baseline решения

Как уже было сказано, проблема baseline решения заключается в том, что модель заведомо не знает, что данные, которые мы пытаемся аппроксимировать обязаны быть неотрицательными. В случае Prophet'a: 1) для линейного тренда - мы никак не можем помешать ему быть отрицательным; 2) для логистического - если даже поставить нижнюю границу 0, компонента сезонности все равно может увести предсказание ниже нуля. Но даже если разобраться с проблемой отрицательных предсказаний модели, с проблемой попадания отрицательных значений в доверительный интервал справится так просто не получится. Но доверительный интервал – ключевой элемент модели для нашей задачи, поэтому очень важно, чтобы помимо предсказания, модель строила подходящий доверительный интервал.

При обучении Prophet [18] максимизирует апостериорное распределение
Правдоподобие для логистического тренда:

$$y \sim \mathcal{N}(C/(1 + \exp(-(k + A \cdot \text{delta}) \cdot (t - (m + A \cdot \text{gamma})))) + X \cdot \text{beta}, \text{epsilon}) \quad (6.11)$$

Правдоподобие для линейного тренда:

$$y \sim \mathcal{N}((k + A \cdot \text{delta}) \cdot t + (m + A \cdot \text{gamma}) + X \cdot \text{beta}, \text{sigma}) \quad (6.12)$$

где $k \sim \mathcal{N}(0, 5)$; $m \sim \mathcal{N}(0, 5)$; $\text{epsilon} \sim \mathcal{N}(0, 0.5)$; $\text{delta} \sim \text{double_exponential}(0, \tau)$; $\text{beta} \sim \mathcal{N}(0, \sigma)$ – априорные распределения. τ, σ – параметры, отвечающие за регуляризацию.

Видим, что в модели неявно предполагается, что данные распределены нормально. И доверительный интервал будет также представлять собой полосу между нижним и верхним квантилями, полученными из параметров апостериорного нормального распределения. Понятно, что с такой реализацией мы не сможем добиться хорошего качества в случае с близкими к нулю значениями.

6.2.1 Внедрение других распределений в реализацию Prophet

В этом пункте мы рассмотрим идею изменения нормального распределения в реализации FBProphet (6.11), (6.12) на какое-нибудь неотрицательное распределение, которое могло бы лучше описать временные ряды, близкие к нулю. Для исследования

были взяты следующие два распределения: гамма и отрицательное биномиальное. Рассмотрим отдельно каждое из них [36].

Гамма-распределение

Определение 6.2.1. Гамма-распределение – это двухпараметрическое семейство абсолютно непрерывных распределений вида

$$\text{Gamma}(x|\alpha, \beta) = x^{\alpha-1} \cdot \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \quad (6.13)$$

где $\alpha, \beta, x \in \mathbb{R}^+$, $\Gamma(\alpha)$ – гамма-функция Эйлера.

Мат. ожидание и дисперсия случайной величины $X \sim \text{Gamma}(\alpha, \beta)$, имеют вид:

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \mathbb{D}(X) = \frac{\alpha}{\beta^2}$$

Данное распределение является неотрицательным, асимметричным, со перекосом вправо: более высокая вероятность встретить редкое “высокое” событие, чем редкое “низкое” событие. Реализация Prophet'а с данным распределением вместо нормального была взята из [37]. Следующие результаты были получены при помощи этой модели.

1. Популярное направление:

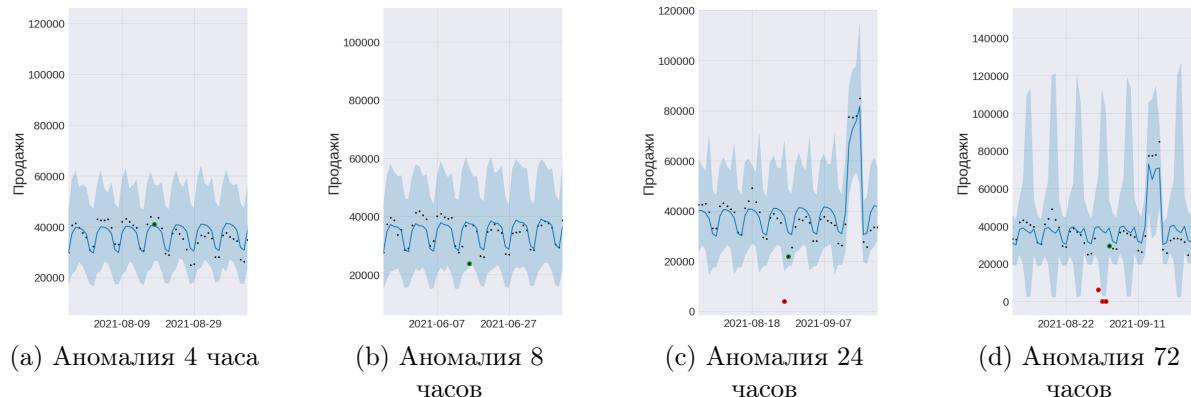


Рис. 6.5: Детекция аномалий моделью Prophet с гамма-распределением на популярном направлении

Таблица 6.4: Метрики Prophet Gamma на популярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.1	0.33	0.99	0.99
Precision	1.0	1.0	1.0	0.99
Recall	0.05	0.2	0.98	1.0

2. Непопулярное направление:

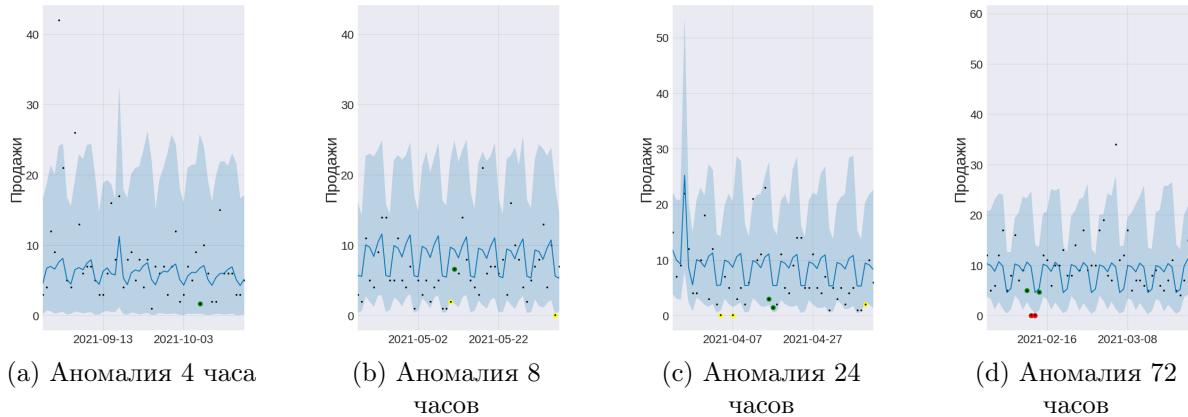


Рис. 6.6: Детекция аномалий моделью Prophet с гамма-распределением на непопулярном направлении

Таблица 6.5: Метрики Prophet Gamma на непопулярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.02	0.01	0.01	0.37
Precision	0.01	0.01	0.05	0.24
Recall	0.08	0.04	0.4	0.32

3. Направление с большим количеством нулевых значений:

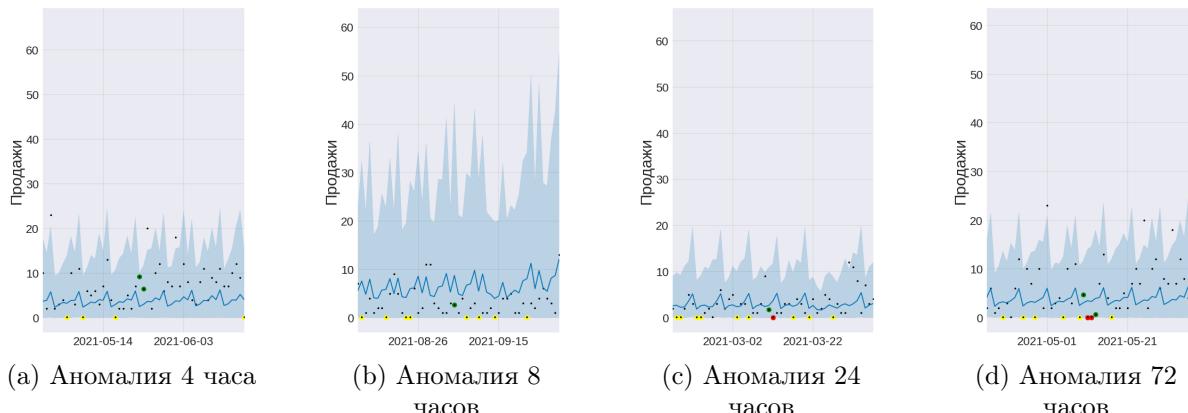


Рис. 6.7: Детекция аномалий моделью Prophet с гамма-распределением на направлении с большим количеством нулей

Таблица 6.6: Метрики Prophet Gamma на направлении с большим количеством нулей

	4 часа	8 часов	24 часа	72 часа
F1	0.01	0.01	0.02	0.1
Precision	0.01	0.01	0.01	0.06
Recall	0.24	0.26	0.04	0.8

Видно, что гамма-распределение гораздо лучше описывает временные ряды с близкими к нулю значениями – доверительный интервал уже не выходит ниже нуля

и даже во многих случаях модели удается отлавливать аномалию длиной 72 часа для близких к нулю временных рядах, о чём говорят и полученные метрики. Также для этого распределения получили близкую к 1 метрику F1 для 24 и 72 часов в случае популярного направления, что тоже является отличным результатом. Из недостатков сразу можно заметить большое количество ложных срабатываний на нулевых продажах и переобучение модели для аномалии длиной 72 часа: на рисунке (6.5d) заметно переобучение сезонной компоненты из-за присутствия 3-х точек сильно ниже ряда, обычный Prophet не имел этой проблемы. Также можно занести в недостатки и разное поведение верхней границы доверительного интервала, как, например, на рисунках (6.7b) и (6.7c) для одного и того же временного ряда модель строит такие разные границы, хотя разница в тренде не такая большая. Эта проблема – следствие асимметрии гамма-распределения. Такую модель из-за особенностей распределения было бы невозможно применять для задачи с верхними аномалиями, но в нашем случае это не принципиальный момент.

Отрицательное биномиальное распределение

Определение 6.2.2. Отрицательное биномиальное распределение – это распределение дискретной случайной величины, равной числу произошедших неудач в последовательности испытаний Бернулли с вероятностью успеха p , проводимых до n -го успеха.

$$\text{NegBinomial}(n|\mu, \phi) = \binom{n + \phi - 1}{n} \left(\frac{\mu}{\mu + \phi}\right)^n \left(\frac{\phi}{\mu + \phi}\right)^\phi \quad (6.14)$$

где $\mu, \phi \in \mathbb{R}^+, n \in \mathbb{N}$.

Мат. ожидание и дисперсия случайной величины $n \sim \text{NegBinomial}(\mu, \phi)$, имеют вид:

$$\mathbb{E}(n) = \mu, \mathbb{D}(n) = \mu - \frac{\mu^2}{\phi}$$

Данное распределение очень похоже по свойствам на гамма-распределение. Существенное отличие в том, что это дискретное распределение. Реализация Prophet'a с данным распределением была взята из [38]. Следующие результаты были получены при помощи этой модели.

1. Популярное направление:

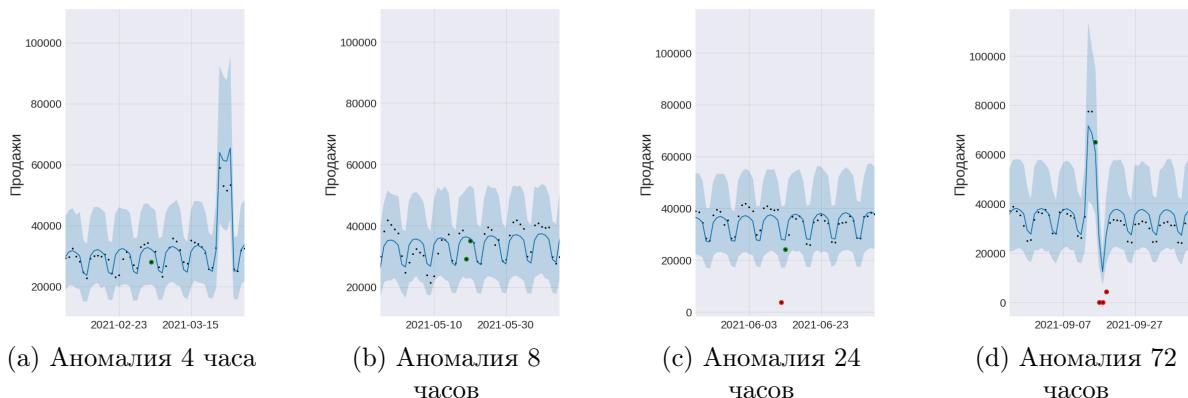


Рис. 6.8: Детекция аномалий моделью Prophet с отрицательным биномиальным распределением на популярном направлении

6.2. ВЫЯВЛЕНИЕ НЕДОСТАТКОВ BASELINE РЕШЕНИЯ ЯРЕШНЕНИЕ ЗАДАЧИ

Таблица 6.7: Метрики Prophet NegBinomial на популярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.1	0.1	0.99	0.99
Precision	1.0	1.0	1.0	1.0
Recall	0.03	0.05	0.98	0.99

2. Непопулярное направление:

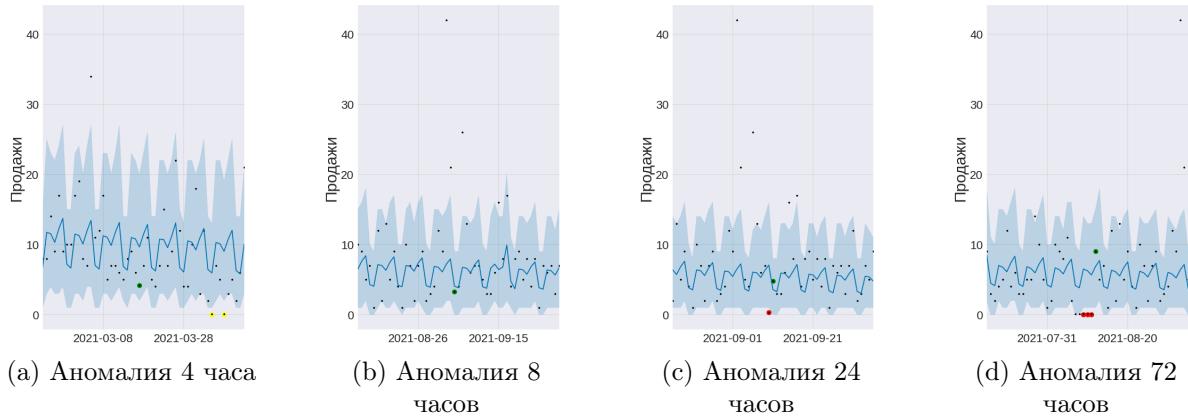


Рис. 6.9: Детекция аномалий моделью Prophet с отрицательным биномиальным распределением на непопулярном направлении

Таблица 6.8: Метрики Prophet NegBinomial на непопулярном направлении

	4 часа	8 часов	24 часа	72 часа
F1	0.02	0.03	0.16	0.44
Precision	0.01	0.02	0.1	0.32
Recall	0.05	0.08	0.5	0.7

3. Направление с большим количеством нулевых значений:

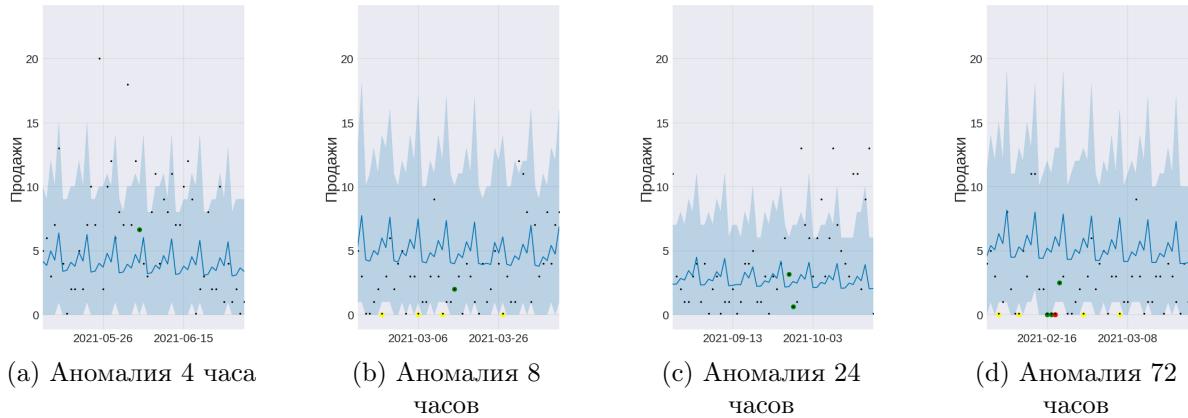


Рис. 6.10: Детекция аномалий моделью Prophet с отрицательным биномиальным распределением на направлении с большим количеством нулей

6.3. ПОДБОР ШИРИНЫ ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА

Таблица 6.9: Метрики Prophet NegBinomial на направлении с большим количеством нулей

	4 часа	8 часов	24 часа	72 часа
F1	0.01	0.01	0.04	0.1
Precision	0.07	0.01	0.02	0.07
Recall	0.09	0.08	0.04	0.34

На первый взгляд, графики похожи на то, что дало гамма-распределение. Но из отличий можно сразу выделить отсутствие переобучения и сильной асимметрии доверительных интервалов. Также из-за того, что отрицательно биномиальное распределение является дискретным, нижняя граница доверительного интервала целочисленна и зачастую совпадает с нулем, из-за чего в таких отрезках ни одна точка не может быть определена, как аномалия. Эта особенность сокращает количество ложных срабатываний, немного уменьшает метрику Recall, но увеличивает Precision, если сравнивать с гамма-реализацией.

6.3 Подбор ширины доверительного интервала

До этого момента мы пытались улучшить только качество предсказания точек временного ряда за счет изменения распределения. Но из формулы (5.1), которую мы используем, как основное определение аномалии для нашей задачи, видно, что мы никак не учитываем один из самых важных параметров задачи τ - ширину доверительного интервала. В этом пункте будет рассмотрена зависимость качества решения от этого параметра.

Как уже уточнялось, Prophet самостоятельно строит доверительный интервал в зависимости от распределения, что очень удобно для нас. Также в модели есть возможность выставить гиперпараметр $\text{interval_width} \in [0, 1]$, как раз отвечающий за ширину доверительного интервала. По умолчанию он равен 0.95 и доверительный интервал в этом случае - полоса между 2.5 % и 97.5 % перцентилями апостериорного распределения, полученного после обучения модели.

Для исследования зависимости метрик от ширины доверительного интервала τ будем, как и до этого, проводить по 80 экспериментов с созданием искусственной аномалии для каждого значения interval_width с 0.6 до 1 включительно с шагом 0.005 (81 значение гиперпараметра). Раньше мы обучали модель 80 (экспериментов) · 4 (интервала аномалии) · 6 (временных рядов) = 1920 раз. Теперь мы должны проделывать то же самое для каждой ширины доверительного интервала и для каждой модели по-отдельности: Prophet (с нормальным распределением), ProphetGamma (с гамма-распределением), ProphetNegBinomial (с отрицательным биномиальным распределением). То есть число моделей для обучения возрастает до $1920 \cdot 81 \cdot 3 = 466\,560$. Это огромное количество моделей получалось обучить и провалидировать за ночь (11-13 часов) на кластере, используя мультипроцессинг и распараллеливая задачу на 36 процессоров.

1. Популярное направление:

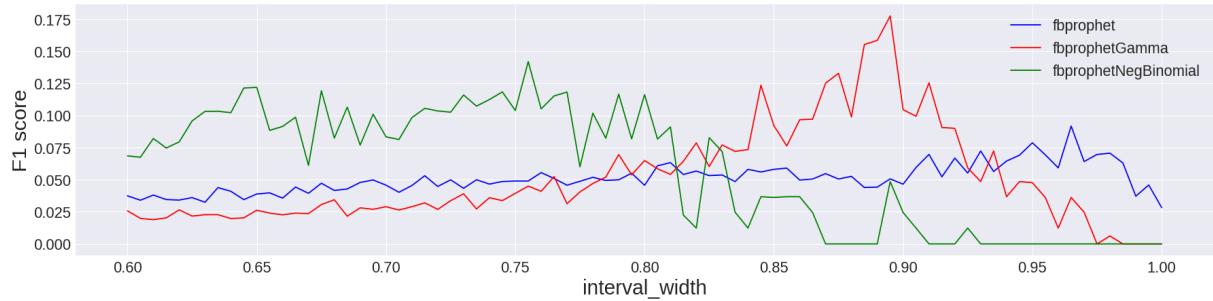


Рис. 6.11: Популярное направление. Аномалия 4 часа

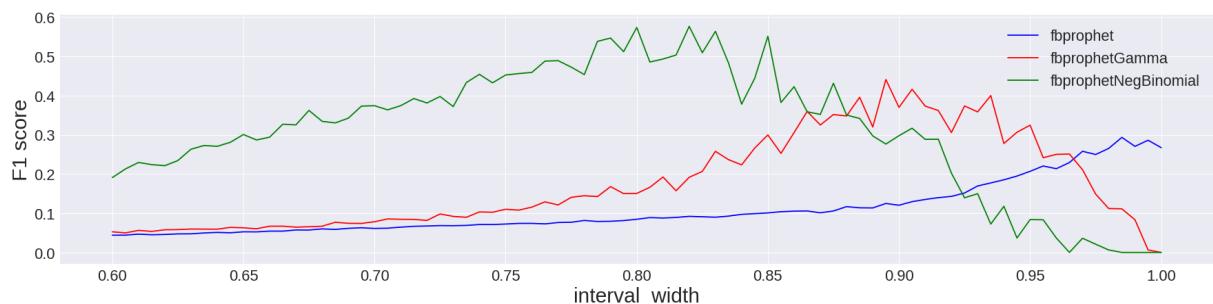


Рис. 6.12: Популярное направление. Аномалия 8 часов

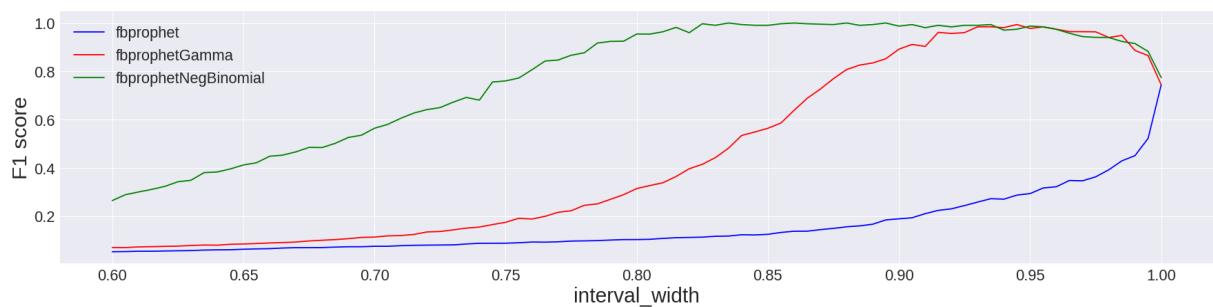


Рис. 6.13: Популярное направление. Аномалия 24 часа

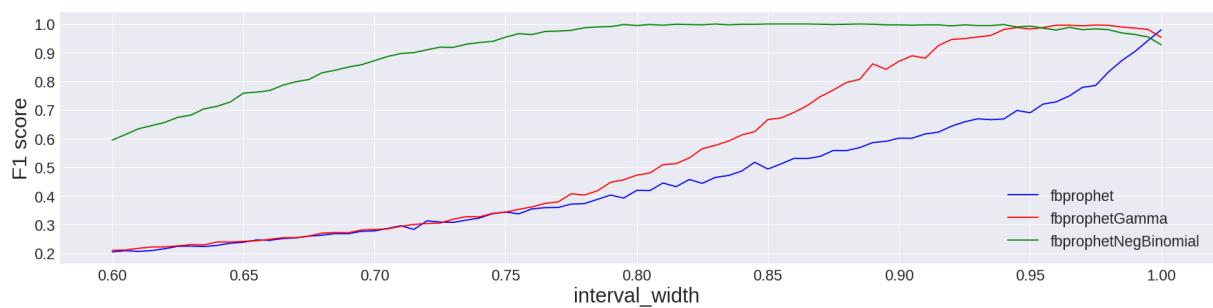


Рис. 6.14: Популярное направление. Аномалия 72 часа

2. Непопулярное направление:

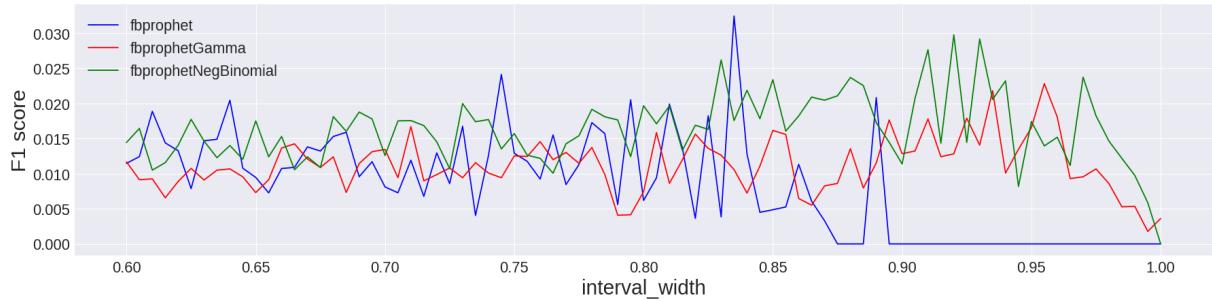


Рис. 6.15: Непопулярное направление. Аномалия 4 часа

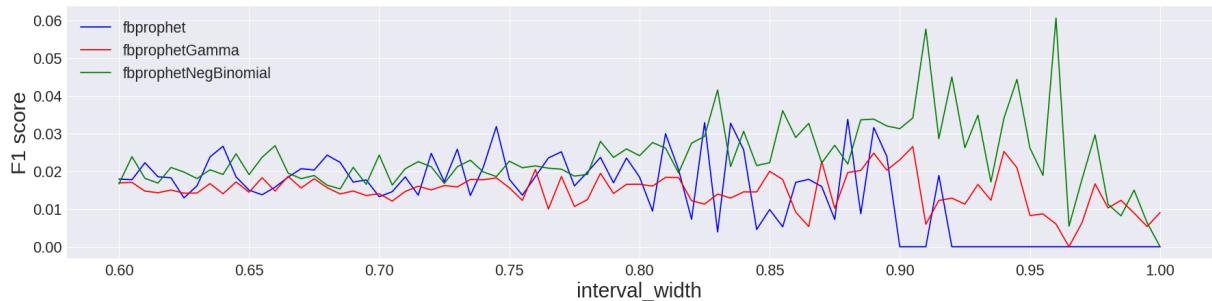


Рис. 6.16: Непопулярное направление. Аномалия 8 часов

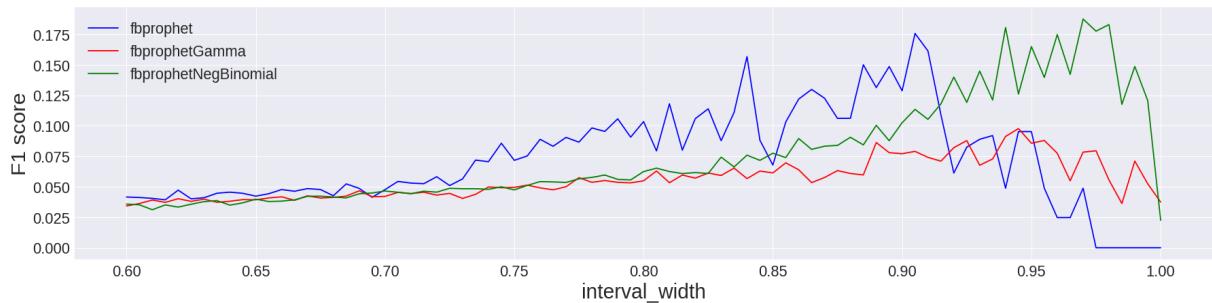


Рис. 6.17: Непопулярное направление. Аномалия 24 часа

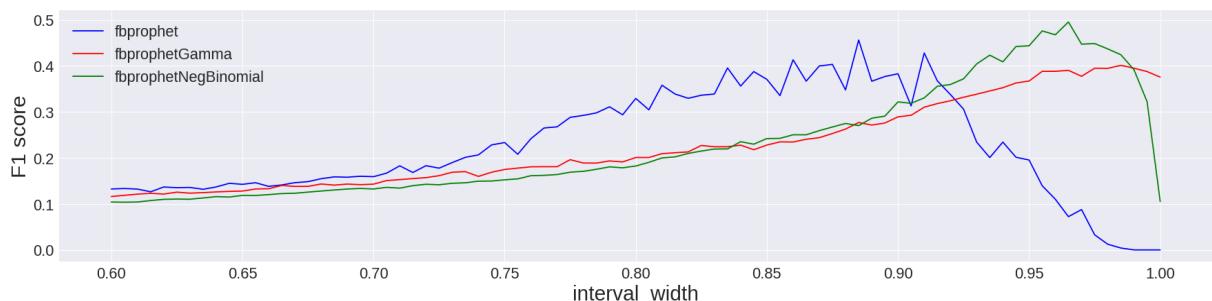


Рис. 6.18: Непопулярное направление. Аномалия 72 часа

3. Направление с большим количеством нулевых значений:

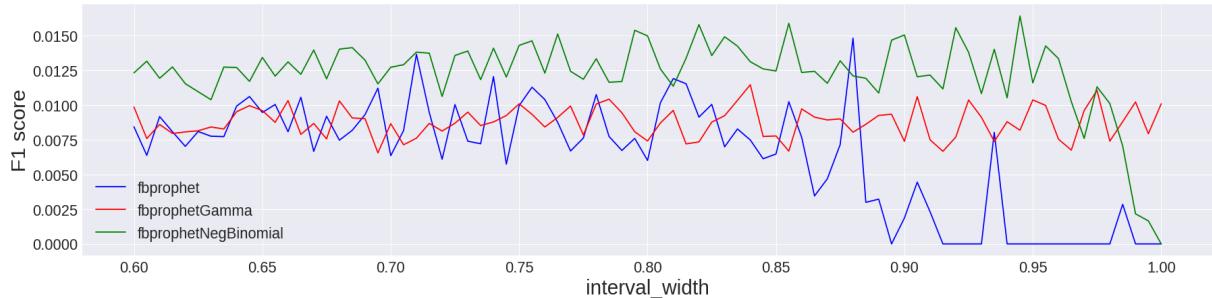


Рис. 6.19: Направление с большим количеством нулевых значений. Аномалия 4 часа

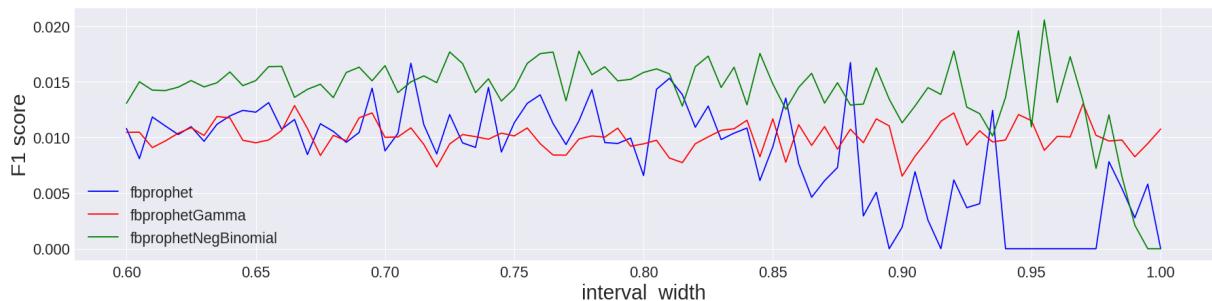


Рис. 6.20: Направление с большим количеством нулевых значений. Аномалия 8 часов

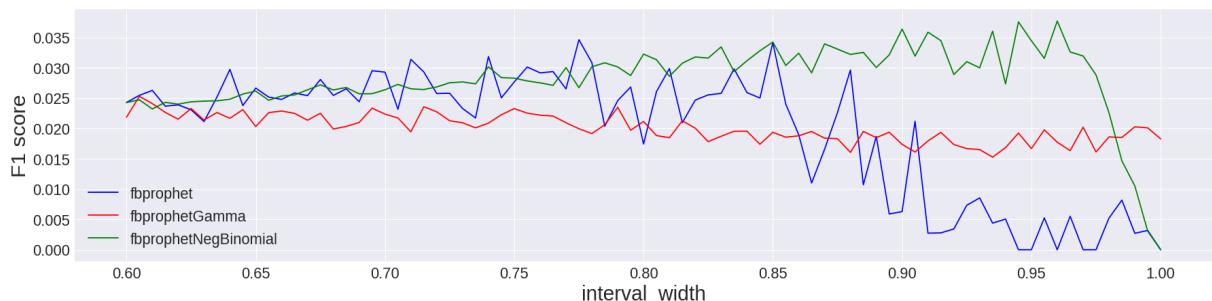


Рис. 6.21: Направление с большим количеством нулевых значений. Аномалия 24 часа

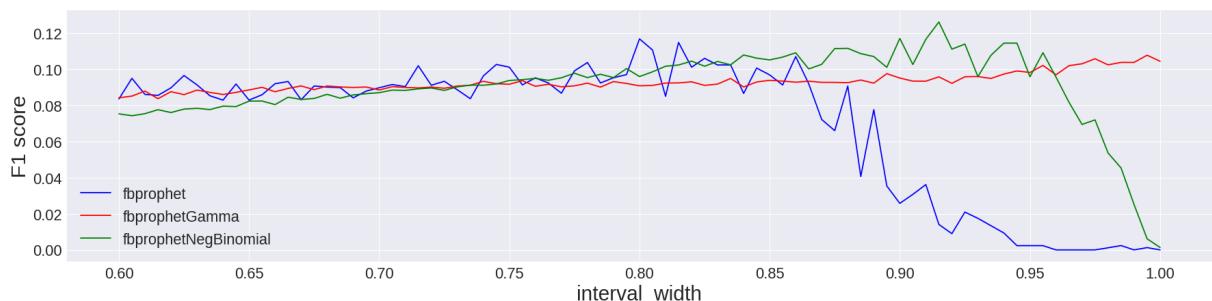


Рис. 6.22: Направление с большим количеством нулевых значений. Аномалия 72 часа

По графикам (6.11) - (6.22) можем увидеть, что для каждого типа направлений и каждого временного интервала аномалии необходимо подбирать свою ширину доверительного интервала для лучшего качества. Данное представление дает наглядное

6.4. ДОБАВЛЕНИЕ СГЛАЖИВАНИЯ ВРЕМЕННОГО РЯДА И ДРАШЕНИЕ ЗАДАЧИ

представление о том, какое качество можно получить для каждого случая в отдельности.

Таблица 6.10: Максимальные метрики F1 при подборе interval_width

	Популярное направ.				Непопулярное направ.				Почти нулевой спрос			
	4 h	8 h	24 h	72 h	4 h	8 h	24 h	72 h	4 h	8 h	24 h	72 h
Prophet	0.1	0.33	0.84	0.99	0.03	0.03	0.18	0.46	0.03	0.05	0.04	0.12
Gamma	0.23	0.49	0.99	0.99	0.02	0.03	0.1	0.41	0.02	0.02	0.03	0.11
NegBin	0.17	0.66	1.0	1.0	0.03	0.06	0.21	0.5	0.02	0.04	0.04	0.13

Из данных результатов видно, что успешно получается решать задачу только для популярного направления и аномалий длиной 8, 24 и 72 часов. Также по графикам явно видно превосходство отрицательного биномиального распределения над другими почти во всех случаях.

Мог возникнуть вопрос: почему не рассматривалась ширина интервала меньше 0.6? Для ответа можем заметить, что чем меньше доверительный интервал, тем больше точек выходят за пределы этого интервала и считаются аномальными, следовательно метрика Recall должна убывать при увеличении interval_width и в нашем случае должна быть максимальной в точке 0.6. Но в этой точке метрика Precision наоборот имеет свой минимум. И, действительно, если посмотреть на конкретные значения, видно, что в данной точке Precision уже близок к нулю, а Recall близок к единице, что уже не удовлетворяет нашим критериям. То есть для меньшей ширины интервала Recall будет стремиться к 1, Precision - к 0, а F1 только уменьшаться. Так что для сокращения числа вычислений было принято решение опустить более мелкие значения interval_width.

6.4 Добавление сглаживания временного ряда

В прошлом пункте мы уже получили улучшение качества наших моделей, но оно все еще не удовлетворяет нашим критериям из пункта (4). В данном разделе мы попробуем улучшить показатели за счет добавления сглаживания данных временного ряда скользящим средним. Различают две разновидности сглаживания: простое сглаживание и взвешенное сглаживание [39].

Определение 6.4.1. *Простое скользящее среднее численно равно среднему арифметическому значений исходной функции за установленный период и вычисляется по формуле*

$$\tilde{X}(k) = \frac{1}{n} \sum_{t=k}^{k+n} X(t) \quad (6.15)$$

где n - размер окна (период сглаживания), k - номер члена ряда, значение которого заменяется средним.

Определение 6.4.2. *Взвешанное скользящее среднее – скользящее среднее, при вычислении которого вес каждого члена исходной функции, начиная с меньшего, равен соответствующему члену арифметической прогрессии.*

$$\tilde{X}(t) = \frac{2}{n(n+1)} \sum_{i=0}^{n-1} (n-i) X(t-i) \quad (6.16)$$

6.4. ДОБАВЛЕНИЕ СГЛАЖИВАНИЯ ВРЕМЕННОГО РЯДА И ДРАШНЕНИЕ ЗАДАЧИ

При вычислении взвешенного скользящего среднего для временного ряда, мы считаем последние значения исходной функции более значимыми чем предыдущие, причем функция значимости линейно убывающая.

В нашем случае все значения являются равнозначными с точки зрения аномальности, поэтому будем использовать простое скользящее среднее.

Данный подход должен помочь в случае с длительной аномалией. Так как длительная аномалия дает нам несколько нулевых продаж подряд, то сглаживание скользящим окном, равным длине аномалии, даст также нулевое значение в центре интервала аномалии. А учитывая то, что в среднем продажи ненулевые или то, что даже в направлении с большим количеством нулей достаточно редко встречаются нулевые продажи несколько дней подряд, скользящее среднее будет давать отличные от нуля значения. Соответственно, эта процедура должна поднимать нулевые значения вверх для обычного сценария и наоборот оставлять у нуля аномальные значения продаж.

Проделаем те же эксперименты для аномалии длительностью 24 и 72 часа, но уже используя сглаживание окнами длиной 2 и 3 дня. Для остальных интервалов аномалий сглаживание скорее всего не даст прироста в качестве, потому что только для данных сценариев интервал больше, чем сутки и усреднение нескольких подряд заниженных показателей будет также заниженным, когда в остальных случаях нельзя точно сказать, как при применении скользящего среднего поведут себя значения ряда и аномалии.

Применять сглаживание скользящим средним можно также несколькими способами:

(1) Сглаживание исходных данных до обучения модели:

Идея данного пункта в том, что для непопулярных направлений факт продаж каждый день зачастую не поддается прогнозированию и является случайной величиной с большой дисперсией. Сглаживание могло бы помочь уменьшить данную дисперсию, что в свою очередь помогло бы модели чуть лучше уловить тренд и отделить аномальные показатели от обычных.

(2) Сглаживание исходных данных после обучения модели:

В прошлом случае есть риск изменить природу данных (тренд и сезонность) за счет сглаживания. В данном же случае такого риска нету, но есть проблема построения доверительного интервала, так как Prophet строит доверительный интервал для случайных величин из исходного распределения, а рассматривать нам предстоит усредненные случайные величины. Чтобы найти их распределения необходимо воспользоваться следующими свойствами.

- $x_1 \in \mathcal{N}(\mu_1, \sigma_1^2), x_2 \in \mathcal{N}(\mu_2, \sigma_2^2)$, то $x_1 + x_2 \in \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Достав все параметры распределений из обученной модели, можно без проблем получить распределение для сглаженных данных, а следовательно и для интервала

- $x_1 \in \text{Gamma}(\alpha_1, \beta), x_2 \in \text{Gamma}(\alpha_2, \beta)$, то $x_1 + x_2 \in \text{Gamma}(\mu_1 + \mu_2, \beta)$

Для СВ из гамма-распределения есть правило сложения только если β у них одинаковые. То есть, чтобы построить новый интервал для усредненных данных, необходимо сделать предположение, что у соседних точек данный параметр не различается.

- $x_1 \in \text{NegBinomial}(n_1, p)$, $x_2 \in \text{NegBinomial}(n_2, p)$,
то $x_1 + x_2 \in \text{NegBinomial}(n_1 + n_2, p)$

Для СВ из отрицательного биномиального аналогично необходимо сделать предположение, что у соседних точек параметр p не различается.

(3) Сглаживание разностей предсказания и исходных данных:

Идея этого подхода состоит в приближении соседних точек, лежащих по разные стороны от прогноза к нему и отдалении точек, лежащих по одну сторону от прогноза, от него. Соответственно, в среднем точки должны сблизиться с прогнозом и только аномальные значения должны оставаться на том же удалении от ожидаемых значений. В данном случае опять встает вопрос с построением доверительного интервала. Но так как мы опять усредняем величины из одного и того же распределения, будем пользоваться методикой из прошлого пункта.

Чтобы не проводить лишних экспериментов, сначала сравним подходы со сглаживанием скользящим окном длиной 2 и 3 дня с подходом без сглаживания и выделим лучшие сценарии применения подходов. А затем сравним между собой все 3 способа сглаживания уже для этих сценариев.

На графиках (6.23) - (6.28) представлено сравнение моделей со сглаживанием (1) типа, длиной окна 2 дня (сплошная линия) и моделей без сглаживания (прерывистая линия). Можно заметить, что сглаживание дает заметный прирост качества только для длины аномалии 72 часа и 2 типов направлений: непопулярного и с большим количеством нулей. Очевидно, что для аномалии длиной 24 часа и меньше окно в 3 дня не даст лучшие результаты, чем окно в 2 дня. Также для популярного направления уже получено максимальное качество и без сглаживания. Поэтому далее, для определения лучшего способа сглаживания будем рассматривать только аномалию 72 часа и 2 типа непопулярных направлений.

1. Популярное направление:

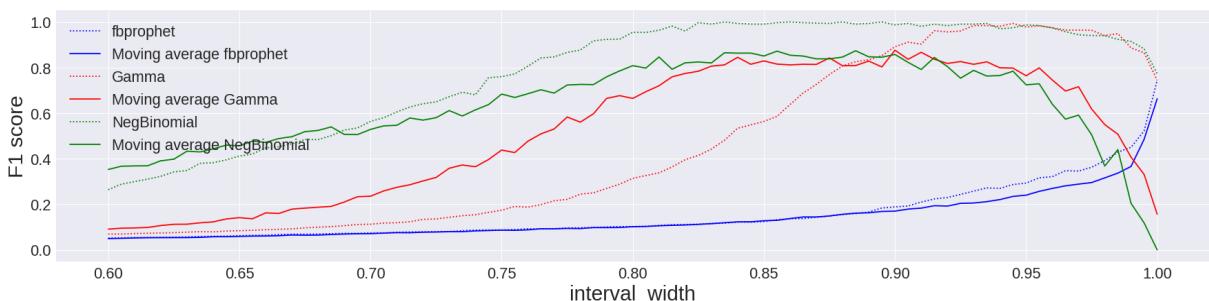


Рис. 6.23: Популярное направление. Аномалия 24 часа

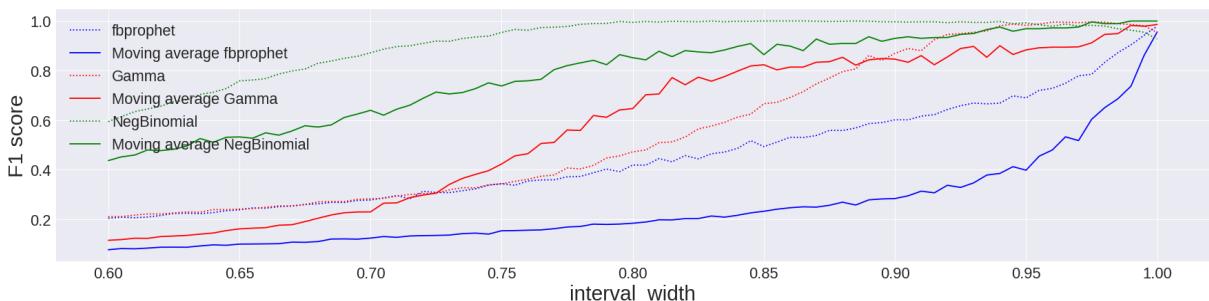


Рис. 6.24: Популярное направление. Аномалия 72 часа

2. Непопулярное направление:

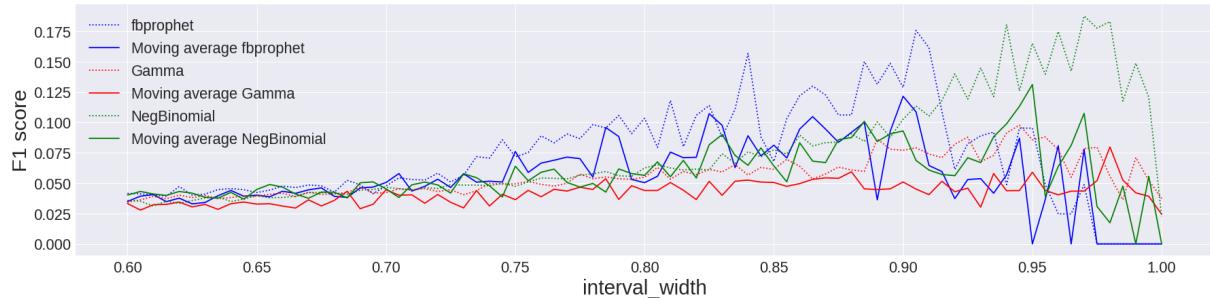


Рис. 6.25: Непопулярное направление. Аномалия 24 часа

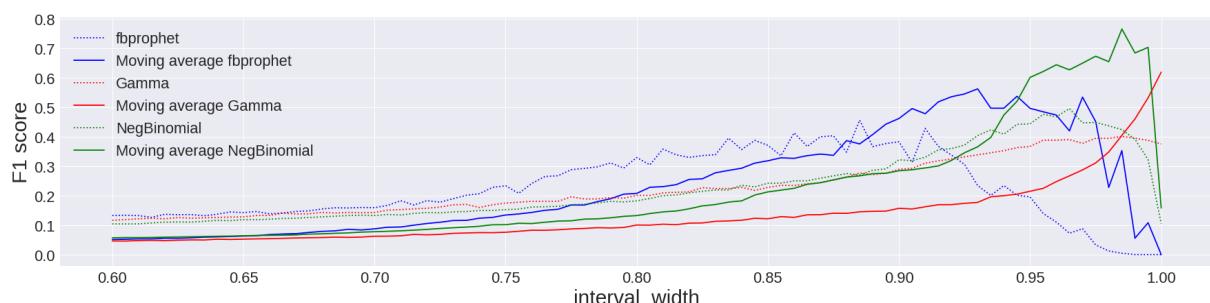


Рис. 6.26: Непопулярное направление. Аномалия 72 часа

3. Направление с большим количеством нулевых значений:

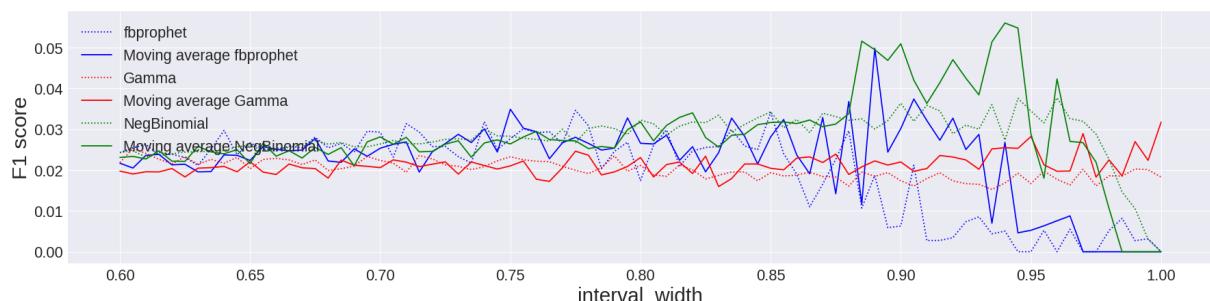


Рис. 6.27: Направление с большим количеством нулевых значений. Аномалия 24 часа

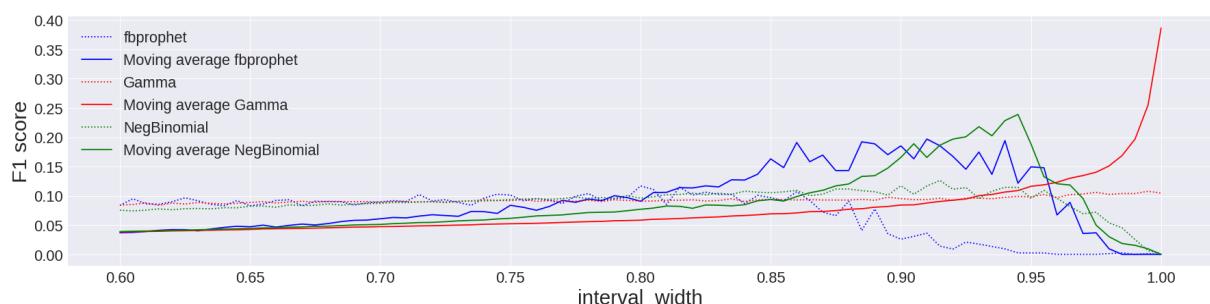


Рис. 6.28: Направление с большим количеством нулевых значений. Аномалия 72 часа

6.4. ДОБАВЛЕНИЕ СГЛАЖИВАНИЯ ВРЕМЕННОГО ПОДАЧИ И ДРАШЕНИЕ ЗАДАЧИ

Таблица 6.11: Сглаживание. Сравнение метрики F1.

		Длина окна = 2 дня		Длина окна = 3 дня	
		Непопулярное напр. 72 h	Почти нулевой спрос 72 h	Непопулярное напр. 72 h	Почти нулевой спрос 72 h
Способ (1)	Prophet	0.56	0.2	0.47	0.22
	Gamma	0.62	0.39	0.74	0.42
	NegBin	0.77	0.24	0.78	0.21
Способ (2)	Prophet	0.45	0.24	0.51	0.25
	Gamma	0.28	0.25	0.52	0.36
	NegBin	0.4	0.23	0.46	0.21
Способ (3)	Prophet	0.64	0.28	0.55	0.24
	Gamma	0.33	0.1	0.37	0.11
	NegBin	0.44	0.15	0.42	0.21

В таблице (6.11) сведены наилучшие (т. е. с подбором оптимальной ширины доверительного интервала `interval_width`) показатели F1 для разных видов сглаживания, моделей и разных типов направлений. Сразу можно увидеть, что в среднем сглаживание окном 3 дня дает лучшие результаты, чем окно в 2 дня. Это связано с тем, что все-таки рассматривается аномалия длиной 72 часа, хоть она и переходящая.

Также можно заметить, что если без сглаживания не получалось успешно решить задачу в случае непопулярного направления, подход сглаживания в некоторых экспериментах дал результаты, удовлетворяющие нашим критериям, что не может не радовать. Хотя для направления с почти нулевым спросом все еще не удалось получить приемлемые метрики.

При сравнении разных способов сглаживания, подход (1) явно выделяется по качеству. То есть можем сделать вывод, что сглаживание исходных данных скользящим окном до обучения модели помогает лучше уловить тренд исходных точек и выявить выбросы. Для (2) и (3) способов сглаживания явно заметно превосходство обычного Prophet'a по сравнению с модификациями. Можно выдвинуть гипотезу, что предположения о сходстве параметров гамма и отрицательно биномиального распределения для соседних точек, которое мы сделали при построении доверительных интервалов в этих способах не совсем корректное и не способствует улучшению результатов.

Глава 7

Анализ полученных результатов

В таблице (7.1) сведены наилучшие результаты для каждого отдельного из рассматриваемых случаев.

Таблица 7.1: Результаты. Метрика F1.

	Популярное направ.				Непопулярное направ.				Почти нулевой спрос			
	4 h	8 h	24 h	72 h	4 h	8 h	24 h	72 h	4 h	8 h	24 h	72 h
Prophet	0.1	0.33	0.84	0.99	0.03	0.03	0.18	0.56	0.03	0.05	0.04	0.25
Gamma	0.23	0.49	0.99	0.99	0.02	0.03	0.1	0.74	0.02	0.02	0.03	0.42
NegBin	0.17	0.66	1.0	1.0	0.03	0.06	0.21	0.78	0.02	0.04	0.04	0.24

Если сравнивать стандартную модель Prophet с ее модификациями, представленными в работе, то можно сказать, что модификации справляются с задачей поиска аномалий лучше, чем базовая модель не только в случае разреженных рядов, но и в случае популярных направлений. Об этом говорят полученные метрики, которые почти во всех случаях выше. При этом базовая модель также позволила получить приемлемые метрики для основных случаев, где было получено успешное решение.

Отдельно можно выделить отрицательное биномиальное распределение, при помощи которого удалось получить единственное решение задачи для аномалии 8 часов на популярных направлениях и максимальные метрики для аномалии 24 и 72 часа на этих же направлениях.

Однако, из графиков пункта (6.2.1) можно сделать вывод, что решение с гамма-распределением подвержено переобучению и иногда нестабильному поведению. Это может быть связано с тем, что для модели с гамма-распределения необходимо выучить в 2 раза больше параметров, чем остальным моделям, что может вносить дополнительную неточность и сложность в решение. Базовый Prophet и модификация с отрицательным биномиальным распределением гораздо лучше аппроксимируют данные и работают более стабильно. Эти рассуждения приводят к заключению, что для решения задачи лучше всего подходит модель с отрицательном биномиальным распределением: одновременно стабильная и дающая лучшее качество.

Таким образом, нам удалось решить поставленную задачу для популярного направления и аномалии длиной 8, 24 и 72 часа и для непопулярного направления, когда сервис продаж не работал на протяжении 72 часов. Для направления с почти нулевым спросом так и не удалось добиться успеха, несмотря на то, что метод сглаживания помог значительно поднять метрики.

Глава 8

Заключение

В данной работе решалась задача поиска точечных аномалий в одномерных временных рядах, в том числе в сильно разреженных рядах, которые представляют собой показатели продаж билетов и доп. услуг с сервисов авиакомпании за каждый день. Данные обладают сильно выраженной сезонностью и интервалами с повышенным спросом во время распродаж.

Для baseline решения этой задачи была выбрана модель Prophet, как модель, которая не требует от временного ряда стационарности (в задаче мы имеем дело с нестационарными рядами), хорошо работает с сезонностью и дополнительными особенностями, как, например, с распродажами, а также предоставляет доверительный интервал, основанный на апостериорном распределении, который можно использовать непосредственно для детекции выбросов.

Но так как baseline модель имеет неявное предположение о нормальном распределении входных данных, она не подходит для описания неотрицательных близких к нулю данных. Для таких случаев были рассмотрены модификации baseline модели с лежащими в их основе гамма и отрицательным биномиальным распределениями, вместо нормального. Данные распределения, являясь положительными, гораздо лучше описывают продажи на непопулярных направлениях.

Для каждой отдельной модели подбиралась оптимальная ширина доверительного интервала, за пределами которого точки считались аномальными. Эксперименты показали, что при использовании модификаций метрика F1 выросла почти для всех направлений и всех длин аномалий. Но Prophet с отрицательным биномиальным распределением ведет себя стабильнее, чем с гамма-распределением, и показывает более высокие метрики качества, поэтому является наиболее эффективным в контексте данной задачи. Применяя эту модель можно со 100% вероятностью находить неисправные сервисы на популярных направлениях, если они не работали больше суток и почти в 70% случаях с допущением 10% ложных срабатываний для тех же направлений, если сервис не работал на протяжении 8 часов.

Также для длительных аномалий был использован подход сглаживания скользящим средним с разной шириной окна. Были исследованы 3 различных способа сглаживания: сглаживание исходных данных до обучения модели, сглаживание исходных данных после обучения модели и сглаживание разностей предсказания и исходных данных. Максимальный прирост к метрикам дало сглаживание с шириной окна равной 3 дня и только для случая аномалии длиной 72 часа, для остальных интервалов аномалий прироста не наблюдалось. По качеству лучшим оказался первый из представленных способов и помог поднять метрики F1 по сравнению с подходом без сглаживания с 0.5 до 0.78 для случая непопулярных направлений и аномалии продолжительностью 72 часа, а также с 0.13 до 0.42 для направлений с почти ну-

левым спросом и аномалии продолжительностью 72 часа. Второй и третий способы сглаживания продемонстрировали чуть худшие метрики, но все же выше, чем модели совсем без сглаживания. Доверительный интервал в этих методах был построен с предположением, что один из параметров распределения соседних точек временного ряда является для них одинаковым. Это предположение близко к реальности, но не истинно. Соответственно, эти способы имеют место на существование, но нуждаются в доработке.

Дальнейшими планами развития этой работы являются следующие шаги:

- Обобщить решение на все имеющиеся ряды

В данной работе были рассмотрены всего 6 различных временных рядов из всех. Необходимо убедиться, что остальные данные по продажам можно разбить на рассмотренные 3 типа направлений, а также, что для временных рядов из одного и того же типа, одни и те же модели, гиперпараметры и способы сглаживания будут давать наилучшие результаты.

- Проанализировать альтернативные модели

Мы исследовали подход с подменой распределения только в модели Prophet, приняв ее за оптимального кандидата для этой задачи по множеству причин, описанных раньше. Вероятно, что найдутся модели, которые лучше подойдут для ее решения и покажут более высокие результаты.

- Внедрить модель в действующую систему детекции

На данный момент в компании время нахождения неисправности в работе сервисов на непопулярных направлениях и направлениях с почти нулевым спросом действующим решением может достигать недели. Внедрение данных подходов поможет уменьшить это время до 3-х дней в большинстве случаев.

Литература

- [1] A review on outlier/anomaly detection in time series data / Ane Blázquez-García, Angel Conde, Usue Mori, Jose A Lozano // *ACM Computing Surveys (CSUR)*. — 2021. — Vol. 54, no. 3. — Pp. 1–33.
- [2] *Basu, Sabyasachi*. Automatic outlier detection for time series: an application to sensor data / Sabyasachi Basu, Martin Meckesheimer // *Knowledge and Information Systems*. — 2007. — Vol. 11, no. 2. — Pp. 137–154.
- [3] Outlier detection in weight time series of connected scales / Saeed Mehrang, Elina Helander, Misha Pavel et al. // 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) / IEEE. — 2015. — Pp. 1489–1496.
- [4] Adaptive threshold for anomaly detection using time series segmentation / Mohamed-Cherif Dani, François-Xavier Jollois, Mohamed Nadif, Cassiano Freixo // International Conference on Neural Information Processing / Springer. — 2015. — Pp. 82–89.
- [5] Automated load curve data cleansing in power systems / Jiyi Chen, Wenyuan Li, Adriel Lau et al. // *IEEE Transactions on Smart Grid*. — 2010. — Vol. 1, no. 2. — Pp. 213–221.
- [6] *Carter, Kevin M*. Probabilistic reasoning for streaming anomaly detection / Kevin M Carter, William W Strelein // 2012 IEEE Statistical Signal Processing Workshop (SSP) / IEEE. — 2012. — Pp. 377–380.
- [7] SCREEN: stream data cleaning under speed constraints / Shaoxu Song, Aoqian Zhang, Jianmin Wang, Philip S Yu // Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. — 2015. — Pp. 827–841.
- [8] *Zhang, Aoqian*. Sequential data cleaning: A statistical approach / Aoqian Zhang, Shaoxu Song, Jianmin Wang // Proceedings of the 2016 International Conference on Management of Data. — 2016. — Pp. 909–924.
- [9] Outlier detection in weight time series of connected scales / Saeed Mehrang, Elina Helander, Misha Pavel et al. // 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) / IEEE. — 2015. — Pp. 1489–1496.
- [10] Using gaussian mixture models to detect outliers in seasonal univariate network traffic / Aarthi Reddy, Meredith Ordway-West, Melissa Lee et al. // 2017 IEEE Security and Privacy Workshops (SPW) / IEEE. — 2017. — Pp. 229–234.
- [11] *Hochenbaum, Jordan*. Automatic anomaly detection in the cloud via statistical learning / Jordan Hochenbaum, Owen S Vallis, Arun Kejariwal // *arXiv preprint arXiv:1704.07706*. — 2017.

- [12] *Akouemo, Hermine N.* Time series outlier detection and imputation / Hermine N Akouemo, Richard J Povinelli // 2014 IEEE PES General Meeting| Conference & Exposition / IEEE. — 2014. — Pp. 1–5.
- [13] *Akouemo, Hermine N.* Probabilistic anomaly detection in natural gas time series data / Hermine N Akouemo, Richard J Povinelli // *International Journal of Forecasting*. — 2016. — Vol. 32, no. 3. — Pp. 948–956.
- [14] *Akouemo, Hermine N.* Data improving in time series using ARX and ANN models / Hermine N Akouemo, Richard J Povinelli // *IEEE Transactions on Power Systems*. — 2017. — Vol. 32, no. 5. — Pp. 3352–3359.
- [15] DeepAnT: A deep learning approach for unsupervised anomaly detection in time series / Mohsin Munir, Shoaib Ahmed Siddiqui, Andreas Dengel, Sheraz Ahmed // *Ieee Access*. — 2018. — Vol. 7. — Pp. 1991–2005.
- [16] *Hill, David J.* Anomaly detection in streaming environmental sensor data: A data-driven modeling approach / David J Hill, Barbara S Minsker // *Environmental Modelling & Software*. — 2010. — Vol. 25, no. 9. — Pp. 1014–1022.
- [17] Statistics-based outlier detection for wireless sensor networks / Yang Zhang, Nicholas AS Hamm, Nirvana Meratnia et al. // *International journal of geographical information science*. — 2012. — Vol. 26, no. 8. — Pp. 1373–1392.
- [18] *Taylor, Sean J.* Forecasting at scale / Sean J Taylor, Benjamin Letham // *The American Statistician*. — 2018. — Vol. 72, no. 1. — Pp. 37–45.
- [19] *Basu, Sabyasachi.* Automatic outlier detection for time series: an application to sensor data / Sabyasachi Basu, Martin Meckesheimer // *Knowledge and Information Systems*. — 2007. — Vol. 11, no. 2. — Pp. 137–154.
- [20] A data quality control method for seafloor observatories: The application of observed time series data in the East China Sea / Yusheng Zhou, Rufu Qin, Huiping Xu et al. // *Sensors*. — 2018. — Vol. 18, no. 8. — P. 2628.
- [21] Anomaly detection in streams with extreme value theory / Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, Christine Largouet // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — 2017. — Pp. 1067–1075.
- [22] A survey on concept drift adaptation / João Gama, Indré Žliobaitė, Albert Bifet et al. // *ACM computing surveys (CSUR)*. — 2014. — Vol. 46, no. 4. — Pp. 1–37.
- [23] *Losing, Viktor.* Incremental on-line learning: A review and comparison of state of the art algorithms / Viktor Losing, Barbara Hammer, Heiko Wersing // *Neurocomputing*. — 2018. — Vol. 275. — Pp. 1261–1274.
- [24] *Xu, Zhao.* Adaptive streaming anomaly analysis / Zhao Xu, Lorenzo von Ritter, Kristian Kersting // Proceedings of NIPS 2016 Workshop on Artificial Intelligence for Data Science. — Vol. 132. — 2016.
- [25] Unsupervised real-time anomaly detection for streaming data / Subutai Ahmad, Alexander Lavin, Scott Purdy, Zuha Agha // *Neurocomputing*. — 2017. — Vol. 262. — Pp. 134–147.

- [26] *Angiulli, Fabrizio.* Distance-based outlier queries in data streams: the novel task and algorithms / Fabrizio Angiulli, Fabio Fassetti // *Data Mining and Knowledge Discovery*. — 2010. — Vol. 20, no. 2. — Pp. 290–324.
- [27] Conformal k -NN Anomaly Detector for Univariate Data Streams / Vladislav Ishimtsev, Alexander Bernstein, Evgeny Burnaev, Ivan Nazarov // Conformal and Probabilistic Prediction and Applications / PMLR. — 2017. — Pp. 213–227.
- [28] *Jagadish, HV.* Mining deviants in a time series database / HV Jagadish, Nick Koudas, S Muthukrishnan // Proceedings of the 25th international conference on very large data bases. — 1999. — Pp. 102–113.
- [29] *Muthukrishnan, Shan.* Mining deviants in time series data streams / Shan Muthukrishnan, Rahul Shah, Jeffrey Scott Vitter // Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004. / IEEE. — 2004. — Pp. 41–50.
- [30] *Croston, John D.* Forecasting and stock control for intermittent demands / John D Croston // *Journal of the Operational Research Society*. — 1972. — Vol. 23, no. 3. — Pp. 289–303.
- [31] *Teunter, Ruud H.* Intermittent demand: Linking forecasting to inventory obsolescence / Ruud H Teunter, Aris A Syntetos, M Zied Babai // *European Journal of Operational Research*. — 2011. — Vol. 214, no. 3. — Pp. 606–615.
- [32] *Nikolopoulos, Konstantinos.* We need to talk about intermittent demand forecasting / Konstantinos Nikolopoulos // *European Journal of Operational Research*. — 2021. — Vol. 291, no. 2. — Pp. 549–559.
- [33] Data-driven Reliability Calculations of Demand Forecast Data / RM Dijkman, IJBF Adan, ZA Bukhsh et al. — 2021.
- [34] Time-series forecasting of seasonal items sales using machine learning—A comparative analysis / Yasaman Ensafi, Saman Hassanzadeh Amin, Guoqing Zhang, Bharat Shah // *International Journal of Information Management Data Insights*. — 2022. — Vol. 2, no. 1. — P. 100058.
- [35] Forecasting intermittent and sparse time series: A unified probabilistic framework via deep renewal processes / Ali Caner Türkmen, Tim Januschowski, Yuyang Wang, Ali Taylan Cemgil // *Plos one*. — 2021. — Vol. 16, no. 11. — P. e0259764.
- [36] *Team, Stan Development.* Stan Functions Reference. — https://mc-stan.org/docs/2_29/functions-reference/index.html.
- [37] Extend model to other distributions 1492. — <https://github.com/facebook/prophet/pull/1492>. — 2020.
- [38] Extend model to other distributions 1492. — <https://github.com/facebook/prophet/pull/1492>. — 2020.
- [39] Moving average. — https://ru.wikipedia.org/wiki/%D0%A1%D0%BA%D0%BE%D0%BB%D1%8C%D0%B7%D1%8F%D1%89%D0%B0%D1%8F_%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D1%8F%D1%8F.

- [40] *Hyndman, Rob J.* Forecasting: principles and practice / Rob J Hyndman, George Athanasopoulos. — OTexts, 2018.
- [41] *Martin, Dominik.* A new metric for lumpy and intermittent demand forecasts: Stock-keeping-oriented prediction error costs / Dominik Martin, Philipp Spitzer, Niklas Kühl // arXiv preprint arXiv:2004.10537. — 2020.